

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:49:27 ; Search time 117 Seconds

(without alignments)  
5507,858 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcagtgagcagagagagcag.....agagtgacctaccagtg 1460

Scoring table: OLIGO\_NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the chance being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	92	6.3	804 4 US-08-932-411A-19	Sequence 19, Appl
2	23	1.6	738 4 US-08-723-570-13	Sequence 13, Appl
3	23	1.6	738 4 US-08-932-411A-13	Sequence 13, Appl
4	23	1.6	1333 1 US-08-910-973-21	Sequence 21, Appl
5	23	1.6	1333 4 US-09-499-227-21	Sequence 21, Appl
6	23	1.6	1385 4 US-08-932-411A-17	Sequence 17, Appl
7	21	1.4	310 1 US-08-552-142A-12	Sequence 12, Appl
8	21	1.4	1268 4 US-08-910-973-12	Sequence 12, Appl
9	21	1.4	1268 4 US-09-499-227-12	Sequence 12, Appl
10	21	1.4	1352 1 US-08-552-142A-10	Sequence 10, Appl
11	21	1.4	1535 1 US-08-910-973-10	Sequence 10, Appl
12	21	1.4	1535 4 US-09-499-227-10	Sequence 10, Appl
13	14	1.4	1550 3 US-09-234-332-3	Sequence 3, Appl
14	13	1.3	50 3 US-08-358-627F-4	Sequence 4, Appl
15	13	1.3	50 3 US-08-793-044-11	Sequence 11, Appl
16	13	1.3	50 4 US-08-465-712C-4	Sequence 4, Appl
17	13	1.3	50 4 US-09-552-733-4	Sequence 4, Appl
18	18	1.2	480 1 US-08-438-123-16	Sequence 16, Appl
19	18	1.2	1336 4 US-09-016-434-129	Sequence 129, Appl
20	18	1.2	1462 1 US-08-552-142A-16	Sequence 16, Appl
21	18	1.2	1951 1 US-08-910-973-16	Sequence 16, Appl
22	18	1.2	1951 4 US-09-499-227-16	Sequence 16, Appl
23	18	1.2	3446 4 US-09-620-312D-653	Sequence 653, Appl
24	18	1.2	4066 4 US-09-367-750-1	Sequence 1, Appl
25	18	1.2	4797 4 US-09-419-568F-25	Sequence 25, Appl
26	18	1.2	4797 4 US-09-354-243B-25	Sequence 25, Appl
27	18	1.2	31728 4 US-09-453-702B-64	Sequence 64, Appl

28	17	1.2	24 3 US-08-358-627F-6	Sequence 6, Appl
29	17	1.2	24 4 US-08-465-712C-6	Sequence 6, Appl
30	17	1.2	24 4 US-09-552-733-6	Sequence 2, Appl
31	17	1.2	24 4 US-09-687-731-2	Sequence 2, Appl
32	17	1.2	24 4 US-09-687-731-8	Sequence 8, Appl
33	17	1.2	24 4 US-09-687-731-10	Sequence 10, Appl
34	17	1.2	510 4 US-09-252-991A-9622	Sequence 9622, Appl
35	17	1.2	685 1 US-08-751-782-5	Sequence 5, Appl
36	17	1.2	685 2 US-08-925-171-5	Sequence 5, Appl
37	17	1.2	846 4 US-09-252-991A-9635	Sequence 9635, Appl
38	17	1.2	948 4 US-09-252-991A-14810	Sequence 14810, Appl
39	17	1.2	954 4 US-09-252-991A-9532	Sequence 9532, Appl
40	17	1.2	1062 4 US-08-086-439C-1	Sequence 1, Appl
41	17	1.2	1161 1 US-08-434-877-1	Sequence 1, Appl
42	17	1.2	1161 1 US-08-434-877-1	Sequence 1, Appl
43	17	1.2	1238 4 US-09-183-861-75	Sequence 75, Appl
44	17	1.2	1238 4 US-09-022-765-75	Sequence 75, Appl
45	17	1.2	1238 4 US-09-551-974A-75	Sequence 75, Appl

## ALIGNMENTS

RESULT 1  
US-08-932-411A-19  
Sequence 19, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr Hobach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 804 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 160..801  
US-08-932-411A-19

**THIS PAGE BLANK (USPTO)**

Query Match 6.3%; Score 92; DB 4; Length 804;  
Best Local Similarity 100.0%; Pred. No. 9.9e-37;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GGGCTGGATGCGGCTGGCGGCTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821  
DB 463 GGGCTGGATGCGGCTGGCGGCTGCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 522  
QY 822 ATGAGACCTTGGGCTTGGCCCACTACAT 853  
DB 523 ATCGAGACCTTGGGCTTGGCCCACTACAT 554

RESULT 2  
US-08-722-570-13

Sequence 13, Application US/08722570  
Patent No. 6555337  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-13

Query Match 1.6%; Score 23; DB 4; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACACT 754  
DB 304 GAGCGCAACCGCATGCACACACT 326

RESULT 3  
US-08-932-411A-13  
Sequence 13, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:  
ADDRESSER: Flehr, Hohbach, Test, Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..732

US-08-932-411A-13

Query Match 1.6%; Score 23; DB 4; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACACT 754  
DB 304 GAGCGCAACCGCATGCACACACT 326

RESULT 4  
US-08-910-973-21  
Sequence 21, Application US/08910973  
Patent No. 5795723  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectode:  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 2, 2004, 15:49:38 ; Search time 3511 Seconds  
(without alignments)  
2493.495 Million cell updates/sec

Title: US-09-595-947E-10  
Perfect score: 1127  
Sequence: 1 MTPQPSGAPVQVTRFERS.....LGATSSACTSPGSLAFSDPL 214

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 segs, 2045481386 residues  
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters: -DEV-xlh  
-MODEL=frame+cpn.model -DEV-xlh  
-O=/cgn2.1/USPTO/spool/US0955947/runat\_02022004\_154933\_8046/app.query.fasta\_1.391  
-DB=genemb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOEXT=0  
-UNITS=bits -START=-1 -END=-1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0955947@cgn.1.1.3508@runat\_02022004\_154933\_8046 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.\*  
1: gb\_da.\*  
2: gb\_htg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_ph.\*  
7: gb\_pl.\*  
8: gb\_pr.\*  
9: gb\_px.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_da.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_scs.\*  
28: em\_un.\*

29: em\_vi.\*  
30: em\_htg\_hum.\*  
31: em\_htg\_inv.\*  
32: em\_htg\_other.\*  
33: em\_htg\_mus.\*  
34: em\_htg\_pln.\*  
35: em\_htg\_rtd.\*  
36: em\_htg\_mam.\*  
37: em\_htg\_vrt.\*  
38: em\_sy.\*  
39: em\_htgo\_hum.\*  
40: em\_htgo\_mus.\*  
41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	100.0	1330	9 HSA133776	AJ133776 Homo sapi
2	1105	98.0	5340	9 AF234829	AF234829 Homo sapi
3	1099	97.5	165110	9 AL450311	AL450311 Human DNA
4	1089	96.6	173341	2 AC021954	AC021954 Homo sapi
5	849	75.3	1491	6 A91167	A91167 Sequence 1
6	849	75.3	1491	6 BD023626	BD023626 Polypepti
7	849	75.3	1491	10 RNRELAXT	Y10619 R.norvegicu
8	849	75.3	258815	2 AC127817	AC127817 Rattus no
9	819	72.7	861	6 AX698801	AX698801 Sequence
10	819	72.7	1861	10 MMU76208	U76208 Mus musculu
11	819	72.7	5567	10 MMU76208	U76207 Mus musculu
12	819	72.7	5567	10 MMU76208	Y99167 M.musculus
13	819	72.7	138070	2 AC109783	AC109783 Mus muscu
14	819	72.7	215050	2 AC127417	AC127417 Mus muscu
15	523	46.4	170896	2 AC011010	AC011010 Homo sapi
16	388.5	34.5	790	5 GGAN012659	AJ012659 Gallus ga
17	388.5	34.5	1880	5 AR303000	AF303000 Gallus ga
18	385.5	34.2	1074	5 GGA012660	AJ012660 Gallus ga
19	382	33.9	1299	5 AF123883	AF123883 Gallus ga
20	381.5	33.9	770	5 AF123884	AF123884 Gallus ga
21	380.5	33.8	1341	5 AF109014	AF109014 Gallus ga
22	378.5	33.6	1268	6 AR023709	AR023709 Sequence
23	378.5	33.6	1268	6 AR225842	AR225842 Sequence
24	378.5	33.6	1675	9 BC008687	BC008687 Homo sapi
25	378.5	33.6	1717	9 BC028226	BC028226 Homo sapi
26	378.5	33.6	79970	2 AC011428	AC011428 Homo sapi
27	378.5	33.6	134506	9 AC005738	AC005738 Homo sapi
28	376.5	33.4	1268	9 HSU03842	U63842 Human neuro
29	375.5	33.3	1385	10 MMU76207	U76207 Mus musculu
30	375.5	33.3	10393	10 AF303001	AF303001 Mus muscu
31	374	33.2	258118	2 AC112007	AC112007 Rattus no
32	374	33.2	258880	2 AC095430	AC095430 Rattus no
33	370.5	32.9	1412	10 MMU76208	Y07621 M.musculus
34	370.5	32.9	151044	2 AC011432	AC011432 Homo sapi
35	370	32.8	1527	6 AR308547	AR308547 Sequence
36	370	32.8	1527	10 RNU67777	U67777 Rattus norv
37	368.5	32.7	123855	2 AC102600	AC102600 Mus muscu
38	365	32.4	932	10 MMU76208	Y09166 M.musculus
39	365	32.4	1315	10 MMU76208	U63841 Mus musculu
40	365	32.4	1333	6 AR023715	AR023715 Sequence
41	365	32.4	1333	6 AR225848	AR225848 Sequence
42	365	32.4	149268	2 AC124395	AC124395 Mus muscu
43	362	32.1	262798	2 AC111702	AC111702 Rattus no
44	359.5	31.9	735	10 MMU67776	U67776 Mus musculu
45	359.5	31.9	738	6 AR308548	AR308548 Sequence

RESULT 1

## ALIGNMENTS

HSA133776  
 LOCUS HSA133776 1330 bp DNA linear PRI 19-JUN-1999  
 DEFINITION Homo sapiens gene for neurogenin 3.  
 ACCESSION AJ133776  
 VERSION AJ133776.1 GI:5123782  
 KEYWORDS bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 AUTHORS Ravassard, P., Icard-Hiepkals, C., Wiard, L., Julien, J.P. and Mallet, J.  
 TITLE The human neurogenin 3 homolog maps to chromosome 10q21.3 and its expression pattern is identical to that of its murine counterparts  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1330)  
 AUTHORS Ravassard, P.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAR-1999) Ravassard P., Igh, CNRS UM69 923, Hopital de la Pitie Salpêtrière, Bat. CERVI, 83 Bd. de l'Hopital, 75013 PARIS, FRANCE  
 FEATURES  
 source  
 1. .1330  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 1. .1330  
 /gene="ngn3"  
 join(<1.157,321. .>1330)  
 /gene="ngn3"  
 1. .157  
 /gene="ngn3"  
 /number=1  
 158. .320  
 /gene="ngn3"  
 /number=1  
 321. .1330  
 /gene="ngn3"  
 /number=2  
 322. .966  
 /gene="ngn3"  
 /function="early neurogenesis"  
 /codon\_start=1  
 /product="neurogenin 3"  
 /protein\_id="CA45384.1"  
 /db\_xref="GI:5123783"  
 /translation="MTGPSGAPVQVTRTERSPRASDEVTCTSPSPPTPTGNCABEAGCGRGAAPRLARGRGRSPKSELALSKQRSRKRRKANDRENRNDNSALDARGVLPDPDDAKLTETLRFAHNYIMALTOTLRADHSLVALPEPPAPHGELGSRGPRGDMGSLYSPSQSGSLSPASLEERRGILGRTSACLSRGSIAFDFL"  
 BASE COUNT 230 a 459 c 413 g 228 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 8.59e-47 Length: 1330  
 Score: 1127.00 Matches: 214  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0  
 US-09-595-947e-10 (1-214) x HSA133776 (1-1330)  
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 322 ATGACGGCTCAACCTCGGGTGCGCCACTGTCGAAGTACCCCGTAGAGCGAGCGGCTC 381  
 QY 21 PheProArgAlaSerGluArgGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 382 TTCCTCCAGAGCTCGGAGACGAGAGTGAAGTCCCGCCCGCCCGCCCGCCCGCCACT 441  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60

DB 442 CCCACACCGGGGAATCTGGCCAGAGCGGACGAGAGGAGGAGCTGCCAGAGGGCCCCAGAGAG 501  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 502 CTCCTCCGGGACCGGCGCGGGGAGCGCAGCCGGCTTGAAGACGAGTGGCCTGAGCAAGCAG 561  
 QY 81 ArgArgSerArgArgLysLysAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 100  
 DB 562 CAGCAGAGTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 621  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120  
 DB 622 TCGGCACTGAGAGCGCTGGCGGTCTCTGCGCACTTCCAGAGAGGAGGAGGAGGAGGAGGAG 681  
 QY 121 LysIleGluThrLeuArgPheAlaHisbentYIleTPAlaLeuThrGlnThrLeuArg 140  
 DB 682 AAGATCGAGACCTCTCGCTCGCCCAACACTACCTGCGGCTGACTCAACAGCTGCGGC 741  
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluIleProProAlaProHisCysGlyGluLeuGly 160  
 DB 742 ATAGCGAGCCACAGCTTACGCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 801  
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180  
 DB 802 AGCCAGAGCGGCTCCCCCGGAGACTGCGGGGTCTCTACTCCCACTCCAGGCTGGC 861  
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuGlyAlaThrSerSer 200  
 DB 862 AGCTGAGTCCCGCGCGCTCGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 921  
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerArgPheLeu 214  
 DB 922 GCTGCTTGAAGCCAGGAGCTGCTGCTTCAAGATTTCG 963  
 RESULT 2  
 AF234829 5340 bp DNA linear PRI 19-OCT-2001  
 LOCUS AF234829 Homo sapiens neurogenin 3 gene, complete cds.  
 DEFINITION AF234829  
 ACCESSION AF234829  
 VERSION AF234829.1 GI:13183002  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 5340)  
 TITLE del Bosque-Plata, L., Lin, J., Horikawa, Y., Schwarz, P.E., Cox, N.J., Iwaseki, N., Ogata, M., Iwamoto, Y., German, M.S. and Bell, G.I.  
 AUTHORS Lin, J. and German, M.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of California San Francisco; 513 Parnassus Ave., San Francisco, CA 94143-0534, USA  
 FEATURES  
 source  
 1. .5340  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 <3022. .>3666  
 /product="neurogenin 3"  
 /codon\_start=1  
 3022. .3666  
 /product="neurogenin 3"  
 /protein\_id="AAK15022.1"  
 /db\_xref="GI:13183003"

```

/translation "MPQPSCAPVQVETERTESPASEDEYTCPTSPAPSPTRTGG
NCAEAEFGCGGACPKULRARGRGRPKSELATSKORSRBRKANDERBRNPHUSG
LDALRGPTPEPDADATKTETLRPAHWYIALVQTLRADSHL, YALEPPACGELG
SPGSGPDWDGSL YSPVSGASLSPLASLEBRPGLLGATSSACLSPGSLAFSDPL"

```

BASE COUNT	1215 a	1500 c	1514 g	1111 f
ORIGIN				

**Alignment Scores:**

Pred. No.:	3,456,45	Length:	5340
Score:	1105.00	Matches:	211
Percent Similarity:	99.078	Conservative:	1
Best Local Similarity:	98.604	Mismatches:	2
Query Match:	98.054	Indels:	0
DB:	9	Gaps:	0

US-09-595-947E-10 (1-214) X AF234829 (1-5340)

QY	1	MethrProGlnProSerGlyValAProThrValGlnValTThrArgGlyLThrGlyAlyrSer	20
Db	3022	ATGAGCCTCAACCTCGGTGTCGCCCACTGTCCAGATGACCCGTGAGACGGAGCGCTCC	3081
QY	21	PheProAlaIAserGluAspGluValThrCysProThrSerAlaProProSerProThr	40
Db	3082	TTCCCCAGAGCCTCGGAGAGCAGAAAGTGAAGCTGCCCAAGTCGCGCCCCCGCCAGCCCACT	3141
QY	41	ArgThrProGlyAAsenCysAlaGluAlaGluGluGlyGlyCysArgGlyValAProArgGlyS	60
Db	3142	CGCACACGGGGGAACTGCGCAGAGGCGGAGAGGGAGGCTGCGAGGGGGCCCCCGAGAGAG	3201
QY	61	LeuArgAlaAArgArgGlyGlyIArgSerArgProIysSerGluLeuAlaLeuSerIysGln	80
Db	3202	CTCCGGGCGACGGCGGGGGGAGCGAGCGCGGCTTAAGACGAGTTGGACTGAGCAAGCAG	3261
QY	81	ArgArgSerAArgArgIysIysAlaAsnAspArgGlyIAspAsnArgMetHisAspLeuAsn	100
Db	3262	CGACGGAATGCGCCAGAAAGAGGCCACAGACCGCAGCGCAATCGAATGACACAACTCAAC	3321
QY	101	SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaIysLeuThr	120
Db	3322	TGGGACATCGAGCGCTCGGGGGTGTCTGGCCACCTTCCGAGACGAGCGAAGCTCAAC	3381
QY	121	LysIleGluThrLeuArgPheAlaHisAsnTyrIleTrrAlaLeuThrGlnThrLeuArg	140
Db	3382	AAGATCGAGAGCGCTCGCTTGGCCCAACATCTGGGGCGTGAATCAACCGCTGGC	3441
QY	141	IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGlyIleuGly	160
Db	3442	ATAGGGAGCCACAGCTTTGACGCGCTGAGACCGCGCGCGCACTGGGGAGAGCTGGGC	3501
QY	161	SerProGlyGlyProProGlyAspTrrGlySerLeuTyrSerProValSerGlnAlaGly	180
Db	3502	AGCCACGAGCGGTTCCTCCGGGGAGCTGGGGGTCTCTCTCCCGCAGTCTCCAGGCTGGC	3561
QY	181	SerLeuSerProAlaAlaSerLeuGluGluArgProGlyIleuLeuGlyValAThrSerSer	200
Db	3562	AGCTGGAATCCCGCGCGCTCGCTGAGAGAGGACCCGGGGCTGTGTGGGGGCCAAGCTTTCC	3621
QY	201	AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu	214
Db	3622	GCTGCTTGAGACCCGAGCAGTGTGGCTTTCACAGATTTTCTG	3683
RESULT 3			
LOCUS	AL450311/c		
DEFINITION	AL450311	165110 bp	DNA linear PRI 12-JUL-2001
ACCESSION	AL450311		Human DNA sequence from clone Rpl1-343J3 on chromosome 10, complete
VERSION	AL450311.11	GI:14626972	
KEYWORDS	HTG.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		

REFERENCE 1 (bases 1 to 165110)  
AUTHORS Howden, P.  
TITLE Direct Submission  
JOURNAL Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL, Sw, SWISSPROT, Tr, TREMBL, Wp, WORMPEP. Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/wormpep/>.

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) this sequence was generated from part of bacterial clone contigs of human Chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch10> RP11-343j3 is from the library RGC1-11.2 constructed by the group of Pister de Jong. For further details see <http://www.choi1.org/bacpac/home.htm>

## FEATURES

## Source

Location/Qu

```
/organism="Homo sapiens"
/mol_type="genomic DNA"
```

/db xref="taxon:9606"

```
/chromosome="10"
```

```
/clone="RP11-343J3"
/clone_lib="PBCT-11 31
```

7 147  
/CJOIE IID- AFCL-11:2

14/

/note="THEIC repeat: matches 2. .142 of consensu

151. .293

```
/note="MIR repeat: matches 2. .148 of consensus"
```

1136. .1187

/note="26 copies 2 mer at 98% conserved"

11025- 20 COPY 2 11025- 20 COPY 2  
0960 3272

[illegible]

/NOTE="ALUSG repeat: matches 1. .310 of consensus

4144. .4270

```
/note="AluX repeat: matches 1. .134 of consensus"
```

4319. .4494

/note="Align repeat: matches 118. .293 of consensus

5663 6438

0002 : 0100  
/0000-011MPC watchdog 077 100E of console

```
repeat_region /note="AluSq repeat: matches 1. .301 of consensus"
14577. .14633
/note="AluSq repeat: matches 243. .299 of consensus"
repeat_region 14843. .14943
/note="MER86 repeat: matches 9. .109 of consensus"
repeat_region 15404. .15613
/note="MIR repeat: matches 3. .213 of consensus"
repeat_region 16626. .16705
/note="MIR repeat: matches 60. .140 of consensus"
repeat_region 16928. .16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region 17476. .17559
/note="MER81 repeat: matches 2. .114 of consensus"
repeat_region 17719. .18069
/note="L1MC4 repeat: matches 7617. .7977 of consensus"
repeat_region 18117. .18345
/note="MIR repeat: matches 7. .262 of consensus"
repeat_region 19169. .19400
/note="MIR repeat: matches 5. .239 of consensus"
repeat_region 19436. .19627
/note="L1MC5 repeat: matches 7728. .7917 of consensus"
repeat_region 19642. .19716
/note="MIR repeat: matches 48. .131 of consensus"
repeat_region 19969. .20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region 20623. .20704
/note="41 copies 2 mer gt 85% conserved"
repeat_region 20738. .20773
/note="U2 repeat: matches 1. .36 of consensus"
repeat_region 20802. .20853
/note="MLT1J repeat: matches 1. .62 of consensus"
repeat_region 20889. .21263
/note="MLT1F repeat: matches 188. .541 of consensus"
repeat_region 21463. .21618
/note="MIR repeat: matches 46. .192 of consensus"
repeat_region 22019. .22326
/note="AluDb repeat: matches 1. .306 of consensus"
repeat_region 22381. .22564
/note="U2 repeat: matches 2453. .2629 of consensus"
repeat_region 22896. .23174
/note="MLT1J repeat: matches 117. .413 of consensus"
repeat_region 23215. .23346
/note="MIR repeat: matches 48. .188 of consensus"
repeat_region 23388. .23532
/note="U2 repeat: matches 2097. .2230 of consensus"
repeat_region 23533. .23837
/note="AluDb repeat: matches 1. .303 of consensus"
repeat_region 23838. .24137
/note="U2 repeat: matches 1754. .2097 of consensus"
repeat_region 24291. .24581
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 24653. .24850
/note="MIR repeat: matches 1. .200 of consensus"
misc_feature 26189. .26685
/note="Cpg Island"
/evidence="not_experimental"
misc_feature 29453. .30918
/note="Cpg Island"
/evidence="not_experimental"
repeat_region 31084. .31156
/note="MIR repeat: matches 65. .138 of consensus"
repeat_region 31618. .31876
/note="AluJo repeat: matches 29. .275 of consensus"
repeat_region 32767. .32830
/note="MIR repeat: matches 76. .139 of consensus"
repeat_region 33050. .33178
/note="43 copies 3 mer tcc 72% conserved"
misc_feature 35112. .36201
/note="Cpg Island"
/evidence="not_experimental"
repeat_region 36551. .36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region 37585. .38254
```

```
repeat_region /note="U1Pn15 repeat: matches 5480. .6157 of consensus"
38265. .38561
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 39214. .39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489. .39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region 40050. .40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057. .41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296. .41460
/note="U2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506. .41944
/note="U2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388. .42698
/note="U2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193. .44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600. .44956
/note="U1R16A repeat: matches 90. .445 of consensus"
repeat_region 45240. .45300
/note="MER8A repeat: matches 37. .97 of consensus"
repeat_region 45798. .45909
/note="U1R41 repeat: matches 90. .192 of consensus"
repeat_region 46826. .46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50684. .50969
/note="AluSq repeat: matches 1. .287 of consensus"
repeat_region 50980. .51291
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 52222. .52519
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 54065. .54260
/note="U1M4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261. .54432
/note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433. .54629
/note="U1M4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648. .54862
/note="U1R41 repeat: matches 11. .217 of consensus"
repeat_region 54863. .55236
/note="U1R1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237. .55700

Alignment Scores:
Pred. No.: 1.42e-43 Length: 165110
Score: 1099.00 Matches: 210
Percent Similarity: 98.60% Conservative: 1
Best Local Similarity: 98.13% Mismatches: 3
Query Match: 97.52% Indels: 0
DB: 9 Gaps: 0

US-09-595-947E-10 (1-214) x AL450311 (1-165110)
QY 1 MethPrpGlnProSerGlyAlaProThrValGlnValThrArgGluThrGlnArgSer 20
DB 30325 ATGAGCGCTCAACCTCGGGTGGCGCCACTGTCCAAAGAGACCGCGAGACGAGCGTCC 30266
QY 21 PheProArgAlaSerGlyAlaPglValThrCysProThrSerAlaProSerProThr 40
DB 30265 TTCCCAAGAGCTCGAAGAGAGAGTGAACGACCGCCACGTCGCCGCCGCCGCCACT 30206
QY 41 ArgThrProGlyAsnGlyAlaGlnValGluGluGlyGlyCysArgGlyAlaProArgGly 60
DB 30205 CGCACACGCGGGGAACTGGCGAGAGCGGAGAGGAGAGGAGCTCCAGAGGCGCCCGAGAG 30146
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGlyLeuAlaLeuSerGlyGln 80
DB 30145 CTCGGGACGCGCGGCGGAGCGAGCGCGCTTAAGACGAGTTGGCACTAGACAGCAG 30086
QY 81 ArgArgSerArgArgGlyGlyAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
```



```
* 53761 59544: contig of 5784 bp in length
* 59545 59644: gap of 100 bp
* 59645 59695: contig of 6751 bp in length
* 66396 66495: gap of 100 bp
* 66496 74696: contig of 8201 bp in length
* 74697 74797: gap of 100 bp
* 81327 81426: contig of 6530 bp in length
* 81427 89446: contig of 8020 bp in length
* 89447 89547: gap of 100 bp
* 89548 96559: contig of 6913 bp in length
* 96560 106498: contig of 9939 bp in length
* 106499 106598: gap of 100 bp
* 106599 116986: contig of 10388 bp in length
* 116987 117086: gap of 100 bp
* 117087 128890: contig of 11804 bp in length
* 128891 128990: gap of 100 bp
* 128991 147290: contig of 18300 bp in length
* 147291 147390: gap of 100 bp
* 147391 173341: contig of 25951 bp in length.
```

FEATURES

```
source 1.173341
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="10"
        /map="10"
        /clone="RP11-57B12"
        /clone_1ib="RP11-11 Human Male BAC"
        1.1574
misc_feature /note="assembly_fragment"
1675..3043
        /note="assembly_fragment"
3144..5889
misc_feature /note="assembly_fragment"
5990..8979
        /note="assembly_fragment"
9080..13674
misc_feature /note="assembly_fragment"
13775..18831
        /note="assembly_fragment"
18932..23526
misc_feature /note="assembly_fragment"
23627..27386
        /note="assembly_fragment"
clone_end:SP6
vector_side:left"
27487..32572
misc_feature /note="assembly_fragment"
32673..38632
        /note="assembly_fragment"
38733..43735
misc_feature /note="assembly_fragment"
43836..49020
        /note="assembly_fragment"
49121..53660
misc_feature /note="assembly_fragment"
53761..59544
        /note="assembly_fragment"
59645..66395
misc_feature /note="assembly_fragment"
66496..74696
        /note="assembly_fragment"
74797..81326
misc_feature /note="assembly_fragment"
81427..89446
        /note="assembly_fragment"
89547..96459
misc_feature /note="assembly_fragment"
96560..106498
        /note="assembly_fragment"
106599..116986
```

```
misc_feature /note="assembly_fragment"
117087..128890
        /note="assembly_fragment"
clone_end:T7
vector_side:right"
misc_feature 128891..147290
        /note="assembly_fragment"
misc_feature 147391..173341
        /note="assembly_fragment"
```

## Alignment Scores:

```
Pred. No.: 4,54e-43 Length: 173341
Score: 1089.00 Matches: 209
Percent Similarity: 98.13% Conservative: 1
Best Local Similarity: 97.66% Mismatches: 4
Query Match: 96.63% Indels: 0
Dbs: 2 Gaps: 0
```

US-09-595-947E-10 (1-214) x AC021954 (1-173341)

```
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 6765 ATGACGCTCAACCTCGGGTGGCGCCCACTGTCGAAGTGAACCCGTGAGACGGAGCGTCC 6706
QY 21 PheProAlaGlnAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 6705 TTCCCAAGAGCTCTCGAAGAGCAAGTGAACCTGCCCAAGTCCGCCCGCCAGCCCACT 6646
QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGluGlyCysArgGlyAlaProArgLys 60
Db 6645 CCGACACGGGGGAGATTGGCCAGAGCGGAGAGGAGAGGAGCTCCAGGGGCCCCGAGAG 6586
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 6585 CTCGGGACACGGCGCGGGGAGACGACCGCGCTTAAGACGAGTTGCACTGAGCAAGCAG 6526
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 6525 CGACGAGTGGCGGAAAGAGCCCAACGACCGCGGCGCAATCGAATCACAACCTCAAC 6466
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db 6465 TCGGCACTGAGCGCTTGCGGCTGCTCTGCCACCTTCCAGACGACGCAAGCTCAAC 6406
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPAlaLeuThrGlnThrLeuArg 140
Db 6405 AAGATCGAAGCGCTCGCTGCGCCCAACTACATCTGGGCGCTGACTCAACGCTGCGC 6346
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 6345 ATAGCGGACCAACAGCTTGACCGCTGAGCGCGCGCGGCGGACCTGGGAGCTGGGC 6286
QY 161 SerProGlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAlaGly 180
Db 6286 AGCCAGGGGGTTCCTCCCGGAGCTGGGGGCTCTCTACTCCCACTCCAGGCTGGC 6226
QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
Db 6225 AGCCTGAGTCCCGCGCGCGTGGAGGAGGAGCAGCCGGGCTGCTGGGGGCGACCTTTCC 6166
QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 6165 GCCTGCTTGAGCCAGGACGTCGTGCTTTCAGATTTCCTG 6124
```

## RESULT 5

```
A91167 A91167 1491 bp DNA linear PAT 22-JAN-2000
LOCUS Sequence 1 from Patent WO9827206.
DEFINITION A91167
ACCESSION A91167
VERSION A91167.1 GI:6740202
KEYWORDS
SOURCE Rattus sp.
ORGANISM Rattus sp.
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 1491)  
 Icard-Jiepkals, C., Mallet, J. and Corresponding, N.A.  
 Patent: WO 9827206-A 1.25-JUN-1998;  
 ICDARD LIBPALMS CHRISTINE (FR); MALLET JACQUES (FR)

FEATURES  
 source  
 1..1491  
 /organism="Rattus sp."  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10118"  
 459..1103  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="GI:6740203"  
 /translation="MAPPLDAPTVQVQSETPQPPASDHEVLSNSTPPSPITVPR  
 DCSSEAGDCRGTSRKLARRGRNRPSELSKQSRKRRKADRBRNMNINSA  
 LDALRGVLPPTDDAKLTETLEPAHYIWMLTQTLRIADHSFYGPPEPVCCEIGS  
 FGGSSGWMGSIVSQAGSLSPASLEEFGLQVPSPSCLPLGLVFDL"

CDS

BASE COUNT 307 a 487 c 413 g 284 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.02e-33 Length: 1491  
 Score: 849.00 Matches: 166  
 Percent Similarity: 83.72% Conservative: 14  
 Best Local Similarity: 77.21% Mismatches: 33  
 Query Match: 75.33% Indels: 2  
 Gaps: 2

US-09-595-947E-10 (1-214) x A91167 (1-1491)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 459 ATGGCGCTCATCTCTTGATGCGCCCAACATCCAAAGTGTCCCAAGAGCAGCAACCC 518  
 QY 21 PheProGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 519 TTTCCCGAGGCTTCGAGCAGCAAGTGTCTCAATTCACCCCACTTACCCCTACT 578  
 QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60  
 DB 579 CTCGACCGAGGAGCTGTCCGAGCAGAGGAGTGTCCGAGGAGCAGCAGAG 638  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 639 CTCCCTGGCGCGCGGAGGCGCAACAGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698  
 QY 81 ArgArgSerArgArgGlyGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 DB 699 CGACCAAGCCGGCGAGAGAGGCAACAGCCGGAGCGCAACCGATGCACAACTTAAAC 758  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 DB 759 TCCCGGCTGAGAGCGCTGCGGTCTCTCCCACTTCCCGAGAGCAGCCAACTTACA 818  
 QY 121 LysIleGluThrLeuArgPheAlaHisAntYrIleTPAlaLeuThrGlnThrLeuArg 140  
 DB 819 AAGATCGAGACCTCGCTTCCGCCCAACTATTTGGGCACTGACTCAGACGCTGGGC 878  
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 DB 879 ATAGCGACCAAGCTTCTACGCCCCCGAGCCCCCTGTGCC--TGTGGGAGCTGGGA 935  
 QY 161 SerPro--GlyGlyProProGlyAspTyrGlySerLeuThrSerProValSerGlnAla 179  
 DB 936 AGCCCGGAGGGGGCTCCAGCGGCGACTGATCTATCTACTCCCAAGTTTCCAAAGCT 995  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 DB 996 GTTACCTGAGGCCCAAGCCCTATTGGAGGAGTTCCCTGGCTCAGGTGCCAGCTCC 1055

QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 1056 CCATCTGTGTGCTCCCGGAGCACCCTGTGTTCTTCAAGCTTCTTG 1100

RESULT 6  
 LOCUS BD023626  
 DEFINITION BD023626 1491 bp DNA linear PAT 27-AUG-2002  
 Polypeptide belonging to the family of basic helix-loop-helix (bHLH) family and nucleic acid sequence corresponding thereto.  
 ACCESSION BD023626.1 GI:22564849  
 VERSION JP 2001510464-A/1.  
 KEYWORDS Rattus sp.  
 SOURCE Rattus sp.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 1491)  
 Iiepkals, C.I., Mallet, J. and Ravassard, P.  
 TITLE Polypeptide belonging to the family of basic helix-loop-helix (bHLH) family and nucleic acid sequence corresponding thereto  
 JOURNAL Patent: JP 2001510464-A 1 31-JUL-2001;  
 RHONE-POULENC RORER SA

COMMENT  
 OS Rattus sp. (rat)  
 PN JP 2001510464-A/1  
 PD 31-JUL-2001  
 PF 19-DEC-1997 JP 1998527415  
 PR 19-DEC-1996 FR 96/15651  
 PI CHRISTINE ICDARD LIBPALMS, JACQUES MALLET, PHILIPPE RAVASSARD PC  
 C07K14/47, A61K31/711, A61K35/76, A61K38/00, A61K48/00, A61P25/00, PC  
 C12N15/09,  
 PC C12N15/00, A61K37/02  
 CC Strandedness: Single;  
 CC Topology: Linear;  
 FH Key

FEATURES  
 source  
 1..1491  
 /organism="Rattus sp."  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10118"

BASE COUNT 307 a 487 c 413 g 284 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 3.02e-33 Length: 1491  
 Score: 849.00 Matches: 166  
 Percent Similarity: 83.72% Conservative: 14  
 Best Local Similarity: 77.21% Mismatches: 33  
 Query Match: 75.33% Indels: 2  
 Gaps: 2

US-09-595-947E-10 (1-214) x BD023626 (1-1491)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 459 ATGGCGCTCATCTCTTGATGCGCCCAACATCCAAAGTGTCCCAAGAGCAGCAACCC 518  
 QY 21 PheProGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 519 TTTCCCGAGGAGCTTCGAGCAGAGGAGTGTCTCAATTCACCCCACTTACCCCTACT 578  
 QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60  
 DB 579 CTCGACCGAGGAGCTGTCCGAGCAGAGGAGTGTCCGAGGAGCAGCAGAG 638  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 639 CTCCCTGGCGCGCGGAGGCGCAACAGCCCAAGAGCGAGTTGGCACTGAGCAAGCAG 698  
 QY 81 ArgArgSerArgArgGlyGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 DB 699 CGACCAAGCCGGCGAGAGGCAACAGCCGGAGCGCAACCGATGCACAACTTAAAC 758



QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLeuThr 120  
 Db TCCGGCTGATGCTGCGGCGGTGCTGCGCCACTTCCGATGAGCCAACTTACA 818  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140  
 Db AAGATCGAGACCTCGCTGCTGCGCCCACTACATTGGGCACTGACGCTGCGC 878  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 Db ATAGGGACCAACAGCTTCTACGCGCCCGACGCCCTGTGCCC---TGTGGGAGCTGGGA 935  
 QY 161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179  
 Db AGCCGGGAGGGGGCTCCAGCGGCGACTGAGGCTCATCTACTCCCAAGTTCCCAAGCT 995  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 Db GGTACCTGAGCCCAACAGCTTCTACGAGAGTTCCCTGGGCTCGAGGTCGCCAGCTCC 1055  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 1056 CCATCTGTCTCTCTCCCGGACCTGCTGTCTCAGACTTCTTG 1100  
 RESULT 7  
 RNRBLAXT 1491 bp mRNA linear ROD 06-MAY-1997  
 LOCUS R. norvegicus mRNA for transcriptional regulator, Relax.  
 ACCESSION Y10619  
 VERSION Y10619.1 GI:2072737  
 KEYWORDS Relax; transcriptional regulator.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE  
 AUTHORS Ravassard, P., Chatail, F., Mallet, D. and Icard-Jilekals, C.  
 TITLE Relax, a novel rat bHLH transcriptional regulator transiently  
 expressed in the ventricular proliferating zone of the developing  
 central nervous system  
 JOURNAL J. Neurosci. Res. 48 (2), 146-158 (1997)  
 MEDLINE 9130143  
 PUBMED 9130143  
 REFERENCE  
 AUTHORS 2 (bases 1 to 1491)  
 TITLE Direct Submission  
 JOURNAL Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,  
 Hopital de la Pitie Salpêtrière, 83 Bd. de l'Hopital, F-75013  
 Paris, FRANCE  
 FEATURES  
 source location/Qualifiers  
 1.1491  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Wistar"  
 /db\_xref="taxon:10116"  
 /clone\_1ib="RT-PCR"  
 459.1103  
 /function="transcriptional regulator"  
 /codon\_start=1  
 /product="Relax"  
 /protein\_id="CAA71630.1"  
 /db\_xref="GI:2072738"  
 /translation="MAHPPLDPTTVOSETQPPFGASDHEVLSNCTPPPTLVR  
 DSEVAGDCRGTSRLKARKGRGNPKSELISKORRRKKANDRENNHNLNSA  
 IDALGVLPDPDADLTETLRFAHNYIMALTOLRLADHSFYGPPEPVCGEIGS  
 PGGSGSGMDGSIYSPVSAQSLSPFASLEEFPLQVPSPPSCILPGLVFSDFL"  
 BASE COUNT 307 a 487 c 413 g 284 t  
 Alignment Scores:

Pred. No.: 3,02e-33 Length: 1491  
 Score: 849.00 Matches: 166  
 Percent Similarity: 83.72% Conservative: 14  
 Best Local Similarity: 77.21% Mismatch: 33  
 Query Match: 75.33% Indels: 2  
 DB: 10 Gaps: 2  
 US-09-595-947E-10 (1-214) x RNRBLAXT (1-1491)  
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db ATGGGCTCTATCCCTTGATGCGCCGACCATTCAGAGTGTCCAGAGACCCAGACACC 518  
 QY 21 PheProAlaGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 Db TTTCCGGAGCTCGAGACAGAGAGTGTCTCAATTCATCCACCCCACTGAGCCCACT 578  
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60  
 Db CTGTAACGAGGAGCTGCTCCAGACAGACAGAGTGTCTCCGAGGAGCATCGAGAGG 638  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db CTCCTGCTGCGCGCGGAGGCGGAGCGGACAGGCCCAAGAGCGAGTTGCACTGAGCAG 698  
 QY 639 CTCCTGCTGCGCGCGGAGGCGGAGCGGACAGGCCCAAGAGCGAGTTGCACTGAGCAG 698  
 Db 639 CGACGAGCCGCGGAGAGAGAGCCAGACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 758  
 QY 81 ArgArgSerArgArgGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 Db 819 AAGATCGAGACCTCGCTGCTGCGCCCACTACATTGGGAGCTGACGACGCTGCGC 878  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLeuThr 120  
 Db 759 TCCGGCTGATGCTGCGGCGGTGCTGCGCCACTTCCGATGAGCCAACTTACA 818  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140  
 Db 819 AAGATCGAGACCTCGCTGCTGCGCCCACTACATTGGGAGCTGACGACGCTGCGC 878  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160  
 Db 879 ATAGGGACCAACAGCTTCTACGCGCCCGACGCCCTGTGCCC---TGTGGGAGCTGGGA 935  
 QY 161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179  
 Db 936 AGCCGGGAGGGGGCTCCAGCGGCGACTGAGGCTCATCTACTCCCAAGTTCCCAAGCT 995  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199  
 Db 936 GGTACCTGAGCCCAACAGCTTCTACGAGAGTTCCCTGGGCTCGAGGTCGCCAGCTCC 1055  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 Db 1056 CCATCTGTCTCTCTCCCGGACCTGCTGTCTCAGACTTCTTG 1100  
 RESULT 8  
 AC127817 258815 bp DNA linear HTG 19-NOV-2002  
 LOCUS Rattus norvegicus clone CH230-259G16, WORKING DRAFT SEQUENCE, 3  
 DEFINITION Rattus norvegicus  
 ACCESSION AC127817 GI:25077905  
 VERSION AC127817.3  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
 SOURCE Rattus norvegicus (Norway rat)  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE  
 AUTHORS 1 (bases 1 to 258815)  
 Muzny, D., Marle, J., Metker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,  
 Allen, C., Allen, H., Albrooks, S., Amin, A., Angilano, D.,  
 Anyalebech, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
 Bissalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
 Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,



Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dedetich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C.M., Gabisi, A., Gamba, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huylk, S., Hume, J., Idelbird, D., Jackson, A., Jacobson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kows, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewe, L., Louised, H., Lozado, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwackelme, O., Okunolu, G., Olarunpungun, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindecker, A., Popovic, D., Primus, B., Fu, L., L., Puafo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmant, K., Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, U., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 258815)  
Worley, K.C.

Direct Submission  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 258815)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 19, 2002 this sequence version replaced gi:23912578. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GZXS  
Center clone name: CH230-259G16  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 224747 bases at least Q40  
Consensus quality: 227981 bases at least Q30  
Consensus quality: 229752 bases at least Q20  
Estimated insert size: 228243; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 255979: contig of 255979 bp in length  
\* 255980 256079: gap of unknown length  
\* 256080 257349: contig of 1270 bp in length  
\* 257350 257449: gap of unknown length  
\* 257450 258815: contig of 1366 bp in length.  
Location/Qualifiers  
1..258815  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-259G16"  
1..1609  
/note="wgs end extension  
clone end:Sp6"  
complement(2865..3659)  
/note="clone boundary  
clone\_end:Sp6  
site:  
end\_sequence:B2169379"  
224098..225157  
/note="clone boundary  
clone\_end:T7  
site:  
end\_sequence:B2169378"  
59351 a 56888 c 58131 g 56821 t 27624 others  
BASE COUNT  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,94e-31 Length: 258815  
Score: 849.00 Matches: 166  
Percent Similarity: 83.72% Conservative: 14  
Best Local Similarity: 77.21% Mismatches: 33  
Query Match: 75.33% Indels: 2  
Gaps: 2  
DB:  
US-09-595-947e-10 (1-214) x AC127817 (1-258815)  
QY 1 MetThrProGlnInProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
Db 247860 ATGGCGCTCATCTCTTGGATGCGCCACCATCAAGTGTCCAGAGACCAAGCAACCC 247919  
QY 21 PheProATGAlaSerGluAlaSPGluValThrCysPProThrSerAlaProProSerProThr 40  
Db 247920 TTTCGGAGGCTCGGACCAAGAGTGTCAATTCCATTCACACCCCACTGAGCCCACT 247979  
QY 41 ArgThrProGlyArgCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60  
Db 247980 CTCGTACGAGGAGCTGCTCGGAGACCAAGAGGTGACTGCGCGAGGAGCATCGAGGAG 248039  
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80

```

Db      248040 CTCCTGCGCGCGCGGAGGCGGCAACAGCCCAAGAGGAGTGGCACTGAGCAAGC 248099
Qy      81  ATGAGSerserArglyLysLysAlaAsnAspArgGluArgAspArgMetHisAspleuAen 100
Db      248100 CGACGAGCGCGCGGCAAGAGGCGCAACGACCGGAGCGCAACCGCATGCAACCTTAA 248159
Qy      101 SerAlaLeuAspAlaLeuArglyValLeuProthrPheProAspAlaLysLeuThr 120
Db      248160 TCCGCGCTGATGAGCTGCGCGGCTGTCGCGCACTTCCCGAGTACCGCAAACTTACA 248219
Qy      121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIetPrAlaLeuThrGlnThrLeuArg 140
Db      248220 AAGATCGAAGCCCTGCGCTTCCGCCCAACATCTTTGGGCACTGACGCGTGGCC 248279
Qy      141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisGlyGlyGluLeuGly 160
Db      248280 ATAGCGGAGCAACAGCTTCTCAACGCGCCCGGAGCCCTCTGTCCTCC---TGTGGAGAGCTGGGA 248336
Qy      161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrlSerProValSerGlnAla 179
Db      248337 AGCCCGGAGGAGGAGGCTCCAGCGGCGGACTGGGCTCTACTACTCCCGAGTTCCCAAGCT 248396
Qy      180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db      248397 GGTACCTGAGGCCCAACAGCTCTATTGAGGAGTTCCCTGCGCTGCAAGTGGCCCAAGCTCC 248456
Qy      200 SerAlaCybLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      248457 CCATCTGTCTCTCTCCCGGAGCACTGTGTCTCTCAAGCTTCTTG 248501

```

```

RESULT 9
LOCUS      AX698801      861 bp      DNA      linear      PART 02-APR-2003
DEFINITION Sequence 7 from Patent WO02086107.
ACCESSION  AX698801
VERSION     AX698801.1 GI:29499589
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  Wobus,A.M., St-Onge,L., Blyszczuk,P. and Hoffmann,U.
            A method for differentiating stem cells into insulin-producing
            cells
JOURNAL     Patent: WO 02086107-A 7 31-OCT-2002;
            Deutscher Aktiengesellschaft fuer Entwicklungsbiologische Forschung
            (DE) ; INSTITUT FÜR PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DE)

```

```

FEATURES
source      1..861
            location/Qualifiers
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            160..804
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAD87442.1"
            /db_xref="GI:29499589"
            /translation="MAHPPLDALTTIQVSEPTQPPFGASDHEVLSSNTPSPPTLIPR
            DCEBAVDCRGSTRLKARGRGRNPKSELISKRRRRKRRKANDRRNRNMINNSA
            IDALRGVLTPTPDADAKLRIETLRPAHNTLMTLTLRADHSFYGPPEPVCGEIGS
            PGGSNGDMGSIYSPSOGKNSPTASLEPFGLOVPSSPYTLRGALVFSDFL"
BASE COUNT  182 a      274 c      250 g      155 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5,32e-32      Length:      861
Score:          819.00      Matches:      163
Percent Similarity: 82.33%      Conservative: 14
Best Local Similarity: 75.81%      Mismatches: 2
Query Match:    72.67%      Indels:      2

```

```

DB:      6      Gaps:      2
US-09-595-947E-10 (1-214) x AX698801 (1-861)
Qy      1  MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db      160 ATGGCGCTCATCTCTTTGGATGCGCTCACCATCAAGTGTCTCCAGAGACACACAACT 219
Qy      21  PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db      220 TTTCCCGAGGCTCGGACACAGAACTGCTCAGTTCCATTCCACCCCACTTGACCCCACT 279
Qy      41  ArgThrProGlyAsnCybAlaGluAlaGluGluGlyGlyCybArgGlyAlaProArgLys 60
Db      280 CTCAATACCTTAGGAGCTGCTCCGAGCAAGATGGGTGACTGCGGAGGAGCTCGAGGAAG 339
Qy      61  LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlyLeuAlaLeuSerLysGln 80
Db      340 CTCGCGCGCCGAGCGGAGGCGGCAACAGCCCAAGAGCGAGTTGGCACTCGACAAACAG 399
Qy      81  ATGAGSerserArglyLysLysAlaAsnAspArgGluArgAspArgMetHisAspleuAen 100
Db      400 CGAAGAGCGCGCGCAAGAGGCGCAATGATCGGAGCGGCAATCGCATGCAACCTTAAC 459
Qy      101 SerAlaLeuAspAlaLeuArglyValLeuProthrPheProAspAlaLysLeuThr 120
Db      460 TCGGCGCTGATGAGCTGCGCGGCTGTCGCTCTCCGAGTGGCCCAAACTTACA 519
Qy      121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIetPrAlaLeuThrGlnThrLeuArg 140
Db      520 AAGATCGAAGCCCTGCGCTTCCGCCCAACATCTTTGGGCACTGACGCGTGGCC 579
Qy      141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisGlyGlyGluLeuGly 160
Db      580 ATAGCGGAGCAACAGCTTCTCAACGCGGAGCCCTGTGTCCTCC---TGTGAGAGCTGGGG 636
Qy      161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrlSerProValSerGlnAla 179
Db      637 AGCCCGGAGGAGGAGGCTCCAGCGGAGCTGAGGCTTACTACTCCAGTCTCCCAAGCG 696
Qy      180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db      697 GGTACCTGAGGCCCAACAGCTCTATTGAGGAGTTCCTGCGCTGCAAGTGGCCCAAGCTCC 756
Qy      200 SerAlaCybLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      757 CCATCTGTCTCTCTCCCGGAGCACTGTGTCTCTCAAGCTTCTTG 801

```

```

RESULT 10
LOCUS      MMU76208      861 bp      DNA      linear      ROD 05-FEB-1997
DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION  U76208
VERSION     U76208.1 GI:1815654
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 861)
            Sommer,L., Ma,Q. and Anderson,D.J.
            neurogenins, a novel family of atonal-related bHLH transcription
            factors, are putative mammalian neuronal determination genes that
            reveal progenitor cell heterogeneity in the developing CNS and PNS
JOURNAL     Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
MEDLINE     97153565
PUBMED      9000438
REFERENCE   2  (bases 1 to 861)
            Sommer,L., Ma,Q. and Anderson,D.J.
            Direct Submission
JOURNAL     Submitted (24-OCT-1996) Biology 216-76, California Institute of
            Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
            USA

```

COMMENT On Feb 5, 1997 this sequence version replaced gi:166911.  
FEATURES Location/Qualifiers

source

1. 861

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

160.804

/gene="ngn3"

160.804

/gene="ngn3"

/function="bHLH transcription factor"

/note="NGN3"

/codon\_start=1

/product="neurogenin 3"

/protein\_id="AAC53029.1"

/db\_xref="GI:166912"

/translation="MAFPLDALTQVSPETQPPRGASDHEVLSNSNTPSPPTLIPR  
DSEAEVGDGRTSRRLARRGRNRPKSELALSKQRSRKRRKADRRNRHNNINSA  
LDALRGVLPFPDDAKLTKEITLRPAHNYIMALTOTLRADHSFYGPPEPVPCGELGS  
PCGSGNGDWGSIYSPVSGAGNISPTASLEBPQLQVPSPSYLLPGLVPSDFL"

BASE COUNT 182 a 274 c 250 g 155 t

ORIGIN

Alignment Scores:

Pred. No.: 5.32e-32 Length: 861  
Score: 819.00 Matches: 163  
Percent Similarity: 82.33% Conservative: 14  
Best Local Similarity: 75.81% Mismatches: 36  
Query Match: 72.67% Indels: 2  
DB: 10 Gaps: 2

US-09-595-947e-10 (1-214) x MMU76208 (1-861)

```

QY      1 MetThrProGlnProSeGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB      160 ATGGGCGCTCATCTCCCTTGAGTGCGCTCACCATCCAAAGTGCTCCAGAGACACAAACACT 219
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB      220 TTTCCCGAGCCTCGAGACCAAGAGTCTCAGTCCCAATTCACCCCACTTACGCCCTACT 279
QY      41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB      280 CTCATCTCAGAGGACTGCTCCAGAGCAAGAGGAGTGAAGTCCGAGGAGCACTCGAGGAAG 339
QY      61 LeuAlaGlaAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB      340 CTCGCGGCCGAGCGCGAGGCGGCAACAGGCCCAAGAGGAGGAGTTCGACCTCAGCAAAACAG 399
QY      81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB      400 CCAAGAGCGCGGCAAGAGGCAAGAGGCAAGTCCGAGGAGGCAAGTCCGAGCAAACTCCAAC 459
QY      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB      460 TCGGGCGCTGAGGCGGTGCGCGGTGCTGCCCACTTCGCGGATGAGCCAAACTTACA 519
QY      121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB      520 AAGATCGAAGACCTGCGCTTCGCGCAACATCACTGCGGCACTGACAGCGTGCAGC 579
QY      141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisValGlyGluLeuGly 160
DB      580 ATAGCGGACCAAGCTTCTATGCGCCGAGGCGCCCTGTGCC---GTGGAAGAGCTGGGG 636
QY      161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuThrSerProValSerGlnAla 179
DB      637 ACGCCCGAGAGTGGCTCCAAAGGAGCTGGGCTCATATCATCTCCCACTTCCCAAGG 696
QY      180 GlySerLeuSerProAlaAsnLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB      697 GGTAACTAGGCGCCCAAGGCTCATATGAGGAATTCCTGCGCTGAGGAGTCCAGCTCC 756

```

QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
DB 757 CCATCTATCTGCTCTCCCGGAGCACTGGTCTTCTAGACTTCTTG 801

RESULT 11.

AF364300

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

```

QY      1 MetThrProGlnProSeGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB      1093 ATGGGCGCTCATCTCCCTTGAGTGCGCTCACCATCCAAAGTGCTCCAGAGACACAAACACT 1152
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB      1153 TTTCCCGAGCCTCGAGACCAAGAGTCTCAGTCCCAATTCACCCCACTTACGCCCTACT 1212
QY      41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB      1213 CTCATCTCAGAGGACTGCTCCAGAGCAAGAGGAGTGAAGTCCGAGGAGCACTCGAGGAAG 1272
QY      61 LeuAlaGlaAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB      1273 CTCGCGGCCGAGCGCGAGGCGGCAACAGGCCCAAGAGGAGGAGTTCGACCTCAGCAAAACAG 1332
QY      81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

```

	/db_xref="taxon:10090"
	/clone="lambda GEM12 clone 19c"
	/cell_line="D3"
	/cell_type="embryonic stem cells"
	/clone_library="genomic library No.62, MboI partial digests cloned in BamHI"
	/dev_stage="embryos"
misc_binding	958
	/bound_moiety="HNF6"
misc_binding	3692
	/bound_moiety="HNF6"
gene	4145..5567
	/gene="MATH4B"
TATA_signal	4145..4151
	/gene="MATH4B"
	/note="putative"
CDS	4923..5567
	/gene="MATH4B"
	/codon_start=1
	/product="neurogenin 3"
	/protein_id="CA470366.1"
	/db_xref="GI:1666088"
	/db_xref="SWISS-PROT:P70661"
	/translation="MAHPDALITIQVSEPTQOPFGASDHVELSNSSTPPSLIPR DCSAGVDGCGTSRKLRARGSGNRKSKSLASKRSRRKANDEMRMHNLNSA LDARGLPTTPPDPAKLTKETELRPANRYMALTOTLRIDHSFGEPEVPQCELSS PQGSNDRKMSISPVSAQNLSPTLASLEFPGLQVSFSYLLRALYSDFL"
BASE COUNT	1271 a 1549 c 1564 g 1183 t.
ORIGIN	
Alignment Scores:	
Pred. No.:	2 79e-31 Length: 5567
Score:	819.00 Matches: 163
Percent Similarity:	82.33% Conservative: 14
Best Local Similarity:	75.81% Miscellaneous: 36
Query Match:	72.67% Indels: 2
DB:	10 Gaps: 2
US-09-595-947E-10 (1-214) x MMATTHAB (1-5567)	
Oy	1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Dd	4923 ATGGGGCTCATCCCTTGATGGCGTCACCATCCAAAGTGTGCCAGAGACACAACAACCT 4982
Oy	21 PheProAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Dd	4983 TTTCGGGAGGCTCGGCACGACGAAGTCTCAGTTCCAATTCCACCACCTAGGCCACT 5042
Oy	41 ArgThrProGlnAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
Dd	5043 CTCATTACTTAGGGAGCTGTCCGAAGCAGAAGTGGGTACTGCCAGGAGACCTCGAGAGAG 5102
Oy	61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Dd	5103 CTCGGCGCGGACGCGGAGGGCGCACACAGGCCAACGAGCGAGTTGGCACTCAGCAACAG 5162
Oy	81 ARGARGSERARGGLYLVALAASNAPRGULARGASNAQMETHIASPLEUASN 100
Dd	5163 CGAAGAAAGCCGCCGACGAAGGCCAAGATCGGAGGGCAGATCCCATGCAACACCTCAAC 5222
Oy	101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Dd	5223 TCGGCGCTGGAGTGGCTGGCGGGGTCTCTGCCACTTCCGGATGAGCCCAACTTACA 5282
Oy	121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIetTrpAlaLeuThrGlnThrLeuArg 140
Dd	5283 AAGATCGAGACCTTCGGCTTCGCCACACTACACTTCGGGACACTGACAGCGCTGGC 5342
Oy	141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Dd	5343 ATAGGGAGCCACAGACTTCTATGAGCCGAGGCCCTGTGGCCC---TGTGGAGAGACTGGGG 5399
Oy	161 SerPro--GlyGlyProProGlnLysPrtGlySerLeuTyrlSerProValSerGlnAla 179

```

Db      5400 AGCCCCGAGGTGGCTCCACCGGAGACTGCTTACTCTCCCAAGTCCCAAGG 5459
Qy      180 GlycerleuserProAlaIalaserleugluGuaArgProGlyleuGluGlyAlaIalaser 199
Db      5460 GGTAACTGAGCCCAACGCGCTTATTGGAGAAATTCCTCGGCTGAGTGCCCAAGCTCC 5519
Qy      200 SerAlaCyLeuserProGlySerleuAlaPheSerAspPheLeu 214
Db      5520 CCATCTATCTGCTCCCGGAGACACTGCTTCTCAGACTTCTTG 5564

RESULT 13
AC109783/c
LOCUS   AC109783
DEFINITION AC109783 138070 bp DNA linear HTG 07-FEB-2002
unpublished pieces.
ACCESSION AC109783.1 GI:18581594
VERSION    HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 138070)
AUTHORS   MacCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimbeno,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zuvever,T.
TITLE      Mouse Genomic Sequence
JOURNAL    Unpublished
REFERENCE 2 (bases 1 to 138070)
AUTHORS   MacCombie,W.R.
TITLE      Direct Submission
JOURNAL    Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT    ----- Genome Center -----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information -----
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Insert size: 141616; sum-of-ctnigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64443: contig of 10669 bp in length
* 64444 74016: gap of unknown length
* 74017 74104: contig of 9573 bp in length
* 74105 83366: gap of unknown length
* 83367 83454: gap of unknown length

```

```

* 83455 92355: contig of 8901 bp in length
* 92356 92443: gap of unknown length
* 92444 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118962: gap of unknown length
* 118962 123619: contig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length
* 128329 132682: contig of 4354 bp in length
* 132683 132770: gap of unknown length
* 132771 136341: contig of 3571 bp in length
* 136342 136429: gap of unknown length
* 136430 138070: contig of 1641 bp in length.
FEATURES
source 1..138070
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="RP23-121F10"
BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Alignment Scores:
Pred. No.: 4,82e-30 Length: 138070
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
Db: Gaps: 2
US-09-595-947E-10 (1-214) x AC109783 (1-138070)
Qy      1 MethPrProGlnProSerGlyAlaProThrAlaGlnValThrArgGluThrGluArgSer 20
Db      111220 ATGGCGCTCATTCCTTGAGATGCGCTCACTCCATCCAGATGTCCTCCAGACACACAACT 111161
Qy      21 PheProAlaIalaserGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db      111160 TTTCCCGAGAGCTCGACACAGAAAGTGTCTCAATTCACACCCCACTAGCCCACT 111101
Qy      41 ArgThrProGlyAsnCyAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
Db      111100 CTCATACCTAGAGACTGCTCCGAGACAGAAAGTGGTACTGCCGAGGAGACTCGAGAA 111041
Qy      61 LeuArgAlaIalaserGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db      111040 CTCGCGCCGACACCGCGGAGGCGGAGAACAGGCCCAAGCGAGTGGCACTCAGCAAAAC 110981
Qy      81 ArgArgSerArgArgLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db      110980 CGAAGAACCGGCGGACAGAGAGCCCAATATGCGAGCCCAATCGCATCACAACCTCAAC 110921
Qy      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Db      110920 TCGGCGCTGAGATGCGCTGCGGAGTGTCTGCGCACTCCCGAGTGAAGCCAACTTACA 110861
Qy      121 LysIleGluThrPheLysArgPheAlaHisAsnArgLysIleThrAlaLeuThrGlnThrLeuArg 140
Db      110860 AAGATCGAGACCTCGCGCTTGGCCCAACAATCGGACATCGACACACGCTGGGC 110801
Qy      141 IleAlaAspHisSerLeuLysArgAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db      110800 ATAAGGACACCACTTCTATAGCCCGGAGCCCTGTGCCC---TGTGGAGAGCTGGGG 110744
Qy      161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuLysSerProValSerGlnAla 179
Db      110743 AGCCCCGAGGTGGCTCCACCGGAGACTGCTTACTCTCCCAAGTCCCAAGG 110684

```



```

Qy 161 SerPro---GlyGlyProProGlyAspTlPGlySerLeuTySerProValSerGlnAla 179
Db 206044 ACCCCGAGGTGGCTCCCAACGGAGCTGGGCTCTATCTACTCTCCCACTCTCCCAAGG 206103
Qy 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 206104 GGTAACCTGAGCCGCCAGCCTCATTTGAGGAATTCCTCGGCTGAGAGTGCACGCTCC 206163
Qy 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAlaPheLeu 214
Db 206164 CCATCTTATCTGCTCCCGGAGACACTGAGTCTCTCAGACTTCTTG 206208

RESULT 15
AC011010/c
LOCUS AC011010
DEFINITION Homo sapiens clone Rpl1-6P16, WORKING DRAFT SEQUENCE, 21 unordered
pieces.
ACCESSION AC011010
VERSION AC011010.4 GI:7107881
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE Birren, B., Linton, L., Nusbaum, C. and Lander, E.
1 (bases 1 to 170896)
JOURNAL Unpublished
2 (bases 1 to 170896)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Balwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dekrellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, J., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galligan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karasik, A., Klein, J.,
Lehoczky, J., Lien, C., Locke, K., Macdonald, P., Margulis, N.,
McEwan, P., McQuirk, J., McKernan, K., McLaughlin, J., Melarini, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessaye, S., Tirrell, A., Vasilev, H., Vo, A., Wheeler, D., Wu, X.,
Wymann, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2000 this sequence version replaced gi:6479051.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information
Project name: L2916
Center project name: 6_P_16
Center clone name: 6_P_16
Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960721
Consensus quality: 114103 bases at least Q40
Consensus quality: 141555 bases at least Q30
Consensus quality: 158230 bases at least Q20
Insert size: 15400; agarose-fp
Insert size: 168896; sum-of-ctrls
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1744: contig of 1744 bp in length
1745 1844: gap of 100 bp
1845 2954: contig of 1110 bp in length
2955 3054: gap of 100 bp
3055 4839: contig of 1785 bp in length
4840 4939: gap of 100 bp
4940 7461: contig of 2522 bp in length
7462 7561: gap of 100 bp
7562 11912: contig of 4351 bp in length
11913 12012: gap of 100 bp
12013 15127: contig of 3115 bp in length
15128 15227: gap of 100 bp
15228 19176: contig of 3949 bp in length
19177 19276: gap of 100 bp
19277 24838: contig of 5562 bp in length
24839 24938: gap of 100 bp
24939 30632: contig of 5694 bp in length
30633 30732: gap of 100 bp
30733 37821: contig of 7089 bp in length
37822 37921: gap of 100 bp
37922 44638: contig of 6717 bp in length
44639 44738: gap of 100 bp
44739 49424: contig of 4686 bp in length
49425 49524: gap of 100 bp
49525 56517: contig of 6993 bp in length
56518 56617: gap of 100 bp
56618 65414: contig of 8796 bp in length
65415 65513: gap of 100 bp
65514 75659: contig of 10146 bp in length
75660 86433: contig of 10674 bp in length
86434 86533: gap of 100 bp
86534 98763: contig of 12230 bp in length
98764 98864: gap of 100 bp
98865 109919: contig of 11056 bp in length
109920 110019: gap of 100 bp
110020 125264: contig of 15245 bp in length
125265 125364: gap of 100 bp
125365 141043: contig of 15679 bp in length
141044 141143: gap of 100 bp
141144 170896: contig of 29753 bp in length.

FEATURES
source
1..170896
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_xref="Rpl1-6P16"
/clone_lib="RPl1-6P16 Human Male BAC"
1..1744
/note="assembly_fragment"
1845..2954
/note="assembly_fragment"
3055..4839
/note="assembly_fragment"
4940..7461
/note="assembly_fragment"
7562..11912
/note="assembly_fragment"
12013..15127
/note="assembly_fragment"
15228..19176
/note="assembly_fragment"
19277..24838
/note="assembly_fragment"
24939..30632
/note="assembly_fragment"
30733..37821
/note="assembly_fragment"

```



```
misc_feature 37922..44638 /note="assembly_fragment"  
misc_feature 44739..49424 /note="assembly_fragment"  
misc_feature 49525..56517 /note="assembly_fragment"  
misc_feature 56618..65413 /note="assembly_fragment"  
misc_feature 65514..75659 /note="assembly_fragment"  
misc_feature 75760..86433 /note="assembly_fragment"  
misc_feature 86534..98763 /note="assembly_fragment"  
misc_feature 98864..109919 /note="assembly_fragment"  
misc_feature 110020..125264 /note="assembly_fragment"  
misc_feature 125365..141043 /note="assembly_fragment"  
misc_feature 141144..170896 /note="assembly_fragment"  
misc_feature 170897..170896 /note="assembly_fragment"  
BASE COUNT 41547 a 41714 c 41241 g 44377 t 2017 others  
ORIGIN
```

```
Alignment Scores:  
Pred. No.: 1,39e-15 Length: 170896  
Score: 523.00 Matches: 99  
Percent Similarity: 98.02% Conservative: 0  
Best Local Similarity: 98.02% Mismatches: 2  
Query Match: 46.41% Indels: 0  
DB: 2 Gaps: 0
```

US-09-595-947E-10 (1-214) x AC011010 (1-170896)

```
QY 114 ProAspAspAlaIysLeuThrIysIleGluThrLeuArgPheAlaHisAsnTyrIleTrp 133  
DB 30631 CCAAGACGACGCGAAGCTCCAGATCGAGACGCTCCGCTCCGCCACAACTACATCTGG 30572  
QY 134 AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAla 153  
DB 30571 GCGCTGACCTCAACGCTGCGCATAGCGGACCAACAGCTGTACGCGCTGAGCGCGCGCG 30512  
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173  
DB 30511 CCGCACTGCGGGAGCTGGGAGCCAGCGGCTTCCCGGGGACTGGGGGCTCCCTTAC 30452  
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193  
DB 30451 TCCCAAGTCTCCAGGCTGGAGCCCTGATCCCGCGCGCTCGTGAGAGAGGACCCGGG 30392  
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySerLeuAlaPheSerAspPhe 213  
DB 30391 CTGCTGGGGGACCTTTTCCGCTGCTTGAAGCCAGGAGCTGCTTCTCAGATTCTT 30332  
QY 214 Leu 214  
DB 30331 CTG 30329
```

Search completed: February 2, 2004, 21:05:31  
Job time : 3780 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 2, 2004, 15:49:38 ; Search time 271 Seconds  
(without alignments)  
2131.660 Million cell updates/sec

Title: US-09-595-947E-10  
Perfect score: 1127  
Sequence: 1 MTPPGSGAPVQVTRTERS.....LCATSSACISPGSLAFSDPL 214

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

```
Command line parameters:
-MODEL=frame+.p2n.model -DEV=xlh
-O=/cgn2.1/USP70_spool/US095594/rnatc_02022004.154933.8036/app_query.faste_1.391
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pico -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US095594@cgn2.1.0 @rnatc_02022004.154933.8036 -NCRU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FCAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

```
Database :
1: N_Geneseq_19Jun03.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1105	98.0	5340	21 AAC61089	Human neurogenin 3
2	1105	98.0	5340	24 AAD46871	Human neurogenin 3
3	849	75.3	1491	19 AAV42512	CDNA encoding a no
4	819	72.7	804	19 AAV27050	Mouse neurogenin 3
5	819	72.7	804	21 AAZ51981	Murine neurogenin-3
6	819	72.7	861	22 AAF27266	Mouse neurogenin-3
7	819	72.7	861	25 ABV75970	Mouse transcrip
8	819	72.7	1860	24 AAD46872	Murine neurogenin
9	819	72.7	1861	21 AAC61090	Murine neurogenin
10	819	72.7	5567	22 AAF27254	Mouse atonal homol
11	519	46.1	592	24 ABQ49522	Oligonucleotide fo
12	519	46.1	592	24 ABQ49523	Oligonucleotide fo
13	467	41.4	592	24 ABQ49524	Oligonucleotide fo
14	467	41.4	592	24 ABQ49525	Oligonucleotide fo
15	388.5	34.5	790	22 AAF27264	Chicken atonal hom
16	385.5	34.2	1074	22 AAF27263	Chicken atonal hom
17	378.5	33.6	1268	18 AAT74891	Human neurogenic d
18	378.5	33.6	1268	19 AAV42932	DNA encoding human
19	378.5	33.6	1268	25 ABS56390	Human bHLH family
20	378.5	33.6	1665	24 AAD46888	Human neurogenin 1
21	376.5	33.4	714	24 AAD46889	Human neurogenin 1
22	375.5	33.3	1385	19 AAV27049	Mouse neurogenin-2
23	375.5	33.3	1385	21 AAZ51980	Murine neurogenin-2
24	375.5	33.3	1385	22 AAF27269	Mouse neurogenin-2
25	370.5	32.9	1412	22 AAF27255	Mouse atonal homol
26	370.5	32.9	1412	22 AAF27273	Mouse atonal homol
27	370	32.8	1527	19 AAV27045	Rat neurogenin-1 g
28	370	32.8	1527	21 AAZ51976	Rat neurogenin-1 g
29	365	32.4	1332	19 AAV42938	DNA encoding murin
30	365	32.4	1332	25 ABS56396	Mouse bHLH family
31	365	32.4	1333	18 AAT74894	Mouse neurogenic d
32	359.5	31.9	738	19 AAV27046	Mouse neurogenin 1
33	359.5	31.9	738	21 AAZ51977	Murine neurogenin-1
34	346	30.7	6123	24 AAD46890	Human neurogenin-2
35	322.5	28.6	1312	19 AAV27047	Xenopus neurogenin
36	322.5	28.6	1312	21 AAZ51978	Xenopus X-ngnr-1a
37	304	27.0	1277	19 AAV27048	Xenopus neurogenin
38	304	27.0	1277	21 AAZ51979	Xenopus X-ngnr-1b
39	297.5	26.4	778	24 ABQ16590	Oligonucleotide fo
40	297.5	26.4	778	24 ABQ16591	Oligonucleotide fo
41	248	22.0	778	24 ABQ16592	Oligonucleotide fo
42	248	22.0	778	24 ABQ16593	Oligonucleotide fo
43	241	21.4	2161	23 TBL13239	Drosophila melanog
44	241	21.4	4161	23 ABL13238	Drosophila melanog
45	238.5	21.2	1550	22 AAF27276	Mouse atonal homol

## ALIGNMENTS

```
RESULT 1
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; chromosome 10g22.1-22.2; cellular differentiation;
KW islet cell precursor identification; diabetes mellitus; human; ds.
XX
OS Homo sapiens.
XX
FH Key
FT CDS Location/Qualifiers
3022..3666
```

```

FT      /*tag= a
FT      /product= "Ngn3"
FT      /note= "Neurogenin 3"
PN      WO200059936-A1.
PD      12-OCT-2000.
XX      .28-MAR-2000; 2000WO-US08436.
XX      06-APR-1999; 99US-0128180.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI, 2000-664889/64.
XX      P-PSDB; AAY85617.
XX      Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX      them, useful for diagnosis, prevention and treatment of diabetes
XX      mellitus and to identify individuals at risk of diabetes -
XX      Claim 6, Page 46-48; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX      protein AAY85617. The Ngn3 gene is located at chromosome position
XX      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX      protein sequences, and includes an antibody recognising the Ngn3 protein.
XX      Also included in the invention is a method for identifying an islet cell
XX      precursor, the method involves analysing a cell for the expression of the
XX      Ngn3 gene product, where detection of the product is indicative of an
XX      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX      reagent for detecting (in a subject) a predisposition to a defect in
XX      pancreatic islet cell function or formation associated with a defect in
XX      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX      precursor cells expressing Ngn3, and to alter cellular differentiation in
XX      culture in vivo to produce new beta-cells to treat patients with diabetes
XX      mellitus.
SQ      Sequence 5340 BP, 1215 A, 1500 C, 1514 G, 1111 T, 0 other;
Alignment Scores:
Pred. No.: 1.65e-59 Length: 5340
Score: 1105.00 Matches: 211
Percent Similarity: 99.07% Conservative: 1
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.05% Indels: 0
DB: 21 Gaps: 0
US-09-595-947E-10 (1-214) x AAC61089 (1-5340)
QY      1 MetThrProGlnProSerGlyAlaProTrpValGlnValThrArgGluThrGluArgSer 20
DB      3022 ATGAGCGCTTCAACCCCTGGGTGGCCCACTGTCACAGTACCCGTTAGACGAGCGCTCC 3081
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProTrp 40
DB      3082 TTCCTCCAGACCTCGGAAGAAGAGACCTGCTCCCACTCGCCCGCCGACCCCTCACT 3141
QY      41 ArgThrProGlyAsnGlyAlaGluValGluGluGlyCysArgGlyAlaProArgGly 60
DB      3142 CGCACACGGGGGAACTGCCAGAGCGGAAGAGGGAGGCTGCCGAGAGGGCCCGAGAGAG 3201
QY      61 LeuAlaGluAlaArgGlyGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerIysGln 80
DB      3202 CTCCTGGGACCGGCGCGGGGACGCAACCGCTTAAGAGCGAGTTGGCACTGACGACAG 3261
QY      81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgGlyHisAspLeuAsn 100
DB      3262 CGACGGAGTCGGGGAAGAGGCAACGCGGACGCGCAATCTGAATGCAACAACCTCAAC 3321
QY      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaIysLeuThr 120

```

```

DB      3322 TCGGCACTGAGACGCCCTCGCGGTCTCTGCCCACTTCCAGACGACGCAAGCTCAC 3381
QY      121 LysIleGluThrLeuAlaArgPheAlaHisAsnTrpIleTrpAlaLeuThrGluThrLeuArg 140
DB      3382 AAGATCGAGACGCTGCGCTTGGCCCACTACATCTGGGCGCTGACTCAACGCTGGCG 3441
QY      141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB      3442 ATAGCGACACACAGCTTGTACCGCTGAGACCGCCGCGCCGCACTGGCGGAGCTGGGC 3501
QY      161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAlaGly 180
DB      3502 AGCCGAGCGGTCTCCCGGGGACTGGGGGTCCTCTACTCCCTCCAGCTGGC 3561
QY      181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB      3562 AGCTGAGTCCCGCGCTGCTGAGGAGGACCCGGGCTGTGGGGGCCACCTCTTCC 3621
QY      201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB      3622 GCTTGTGAGCCCAAGGAGTCTGCTTCTCAATTTTCTG 3663
RESULT 2
AAD46871
ID      AAD46871 standard; DNA; 5340 BP.
XX
AC      AAD46871;
XX
DT      27-JAN-2003 (first entry)
XX
DE      Human neurogenin 3 (Ngn3) gene.
XX
KW      Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW      type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW      islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
KW      gene; ds.
XX
OS      Homo sapiens.
XX
FH      Key Location/Qualifiers
FT      CDS 3022..3666
FT      FT /*tag= a
FT      /product= "Human Ngn3 protein"
PN      WO200274045-A2.
XX
PD      26-SEP-2002.
XX
PF      20-MAR-2002; 2002WO-US11166.
XX
PR      20-MAR-2001; 2001US-0817360.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI, 2002-759853/82.
XX      P-PSDB; AAE29277.
XX
PT      Producing a mammalian islet cell for treating diabetes mellitus
PT      comprises introducing into a mammalian cell a nucleic acid molecule
PT      encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
PS      Example 2, Page 87-88; 108pp; English.
XX
CC      The invention relates to a method for producing a mammalian islet cell.
CC      The method comprising introducing into a mammalian cell a nucleic acid
CC      molecule encoding an islet transcription factor for expression of the
CC      islet transcription factor in the cell and for production of islet cell
CC      phenotype in the cell. The islet transcription factor is a neuroendocrine
CC      basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC      for treating type 2 diabetes mellitus and for replacing beta cells lost

```

CC to autoimmune destruction in individuals with type 1 diabetes. The method  
CC is useful in cell therapy. The present sequence is human neurogenin 3  
CC (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.

XX Sequence 5340 BP, 1215 A, 1500 C, 1514 G, 1111 T, 0 other;

#### Alignment Scores:

Pred. No.:	1-65e-59	Length:	5340
Score:	1105.00	Matches:	211
Percent Similarity:	99.07%	Conservative:	1
Best Local Similarity:	98.60%	Mismatches:	2
Query Match:	98.05%	Indels:	0
DB:	24	Gaps:	0

US-09-595-947E-10 (1-214) x AAD6871 (1-5340)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 3022 ATGACGCGCTCAACCTCGCGGTGCGCCCACTGTCGAAGTACCCTGAGACGAGCGGTCC 3081
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 3082 TTCCTCCAGAGCTCGAGAACAGAGTGAAGTCCCTCCAGCTCCCGCCCGCCCACT 3141
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 3142 CGCACACGGGGGAACTGCCAGAGCGGAGAGGAGGCTGCCGAGGGGCCCCGAGGAAG 3201
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 3202 CTCGGGCGACGGCGGGGAGAGCGAGCCGCGCTAAGAGCGAAGTGGTCATGAGCAAGCAG 3261
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 3262 CCACGAGCTGGCGGAAAGAGCCCAAGCAGCCGAGCGGATCGAATGACAACTCAAC 3321
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 3322 TCGGCACTGAGCGCTGCGCGGTGCTCTGCCCACTTCCAGACGAGCGAGCTCACC 3381
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 3382 AAGATCGAGACGCTGCGCTTCCGCCCACTACATCTGGCGCTGACTCAAGCGTGGCC 3441
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 3442 ATAGCGGACACAGCTTGTACCGCGCTGAGCGCGCCCGCACCTGCGGAGCTGGGC 3501
QY 161 SerProGlyGlyProProGlyAspTyrPheLysSerLeuTyrSerProValSerGlnAlaGly 180
DB 3502 AGCCGAGCGGCTTCCCGCGGAGCTGGGGGTCCCTTACTTCCCACTTCCCAAGCTGGC 3561
QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB 3562 ACGCTGAGTCCGCGCGCTGCTGAGAGAGCGAGCCCGGCGTGGGCGGCGACCTCTTCC 3621
QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 3622 GCGTCTGAGCCAGGAGCTGCGCTTCTCAGATTCTTG 3663

```

#### RESULT 3

AAV42512 standard; cDNA, 1491 BP.

AAV42512;

05-OCT-1998 (first entry)

CDNA encoding a novel BHLH protein designated RELAX.

Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;

KW control; gene expression; transcriptional activator; targeting;

KW protein expression; central nervous system; CNS; treatment;

KW nervous system disorder; se.

XX Rattus sp.

XX OS

XX Key

XX CDS

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

XX FT

Location/Qualifiers  
459..1103  
/\*tag= a  
/product= RELAX

W09827206-A2.

25-JUN-1998.

19-DEC-1997; 97WO-FR02368.

19-DEC-1996; 96FR-0015651.

(RHON ) RHONE-POULENC RORER SA.

Mallet J, Ravassard P, Icard-Liepkalns C;

WPI, 1998-36275/31.

P-PSDB; AAM62991.

Basic helix-loop-helix polypeptide and related nucleic acid - with  
transcriptional activity, for targeting expression of genes to  
central nervous system and treatment of nervous disease

Claim 6, Page 20; 28pp, French.

The present sequence encodes a basic helix-loop-helix (BHLH) type  
protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.  
The protein is used to control and participate in gene expression,  
by acting as transcriptional activator, strictly dependent on the  
presence of an intact E box (CANNTG), particularly for targeting  
expression of proteins to the central nervous system (CNS). The  
CC nucleic acid sequence can be used to treat nervous system disorders,  
and antisense sequences can be used to control mRNA transcription.

Sequence 1491 BP, 307 A, 487 C, 413 G, 284 T, 0 other;

#### Alignment Scores:

Pred. No.:	3-5e-44	Length:	1491
Score:	849.00 <td>Matches:</td> <td>166</td>	Matches:	166
Percent Similarity:	83.72%	Conservative:	14
Best Local Similarity:	77.21%	Mismatches:	33
Query Match:	75.31%	Indels:	2
DB:	19	Gaps:	2

US-09-595-947E-10 (1-214) x AAV42512 (1-1491)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCTCATCTCCCTTGATGCGCCCACTCCAAAGTGTCCCAAGAGACGACGACACC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTCCTCCGAGCTCGAGACAGAGTGTCAATTCCACCCCACTTACCTTACCTTACCTTAC 578
QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTAACGAGGAGTGTCTCCGAGACAGAGAGTGACTGCGAGGAGCATCGAGGAAG 638
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCCTGCGCGCGCGGAGGCGGCAAGGCGCCCAAGACCGAGTGGCACTGAGCAAGCAG 698
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 699 CGACGAGCGCGGAGAGAGCCCAAGCAGGAGCGGACCGAGTGCACAACTTAC 758
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

```

Db 759 TCAGCGCTGATGAGCTGCGGAGTGTCTGCCACCTCCGAGANGACCAAACTTACA 818  
 Qy 121 LysIleGIuThrLeuAArgPheAlaHisAntYrIleTPrAlaLeuThrGlnThrLeuArg 140  
 Db 819 AAGATCGAAGACCTGCGCTTCCGCCACACTACATTGGACACTGACAGCGTCCGC 878  
 Qy 141 IleAlaAPHisSerLeuTyrAlaLeuGIuProProlAProHicCySGlyGIuLeuGIy 160  
 Db 879 ATAGCGGACCAAGCTTCAACGCGCCGAGCCCTGTCCTCC---TGTGGGAGCTGGGA 935  
 Qy 161 SerPro---GlyGIuProProGIuAsPTrGIuSerLeuTyrSerProValSerGlnAla 179  
 Db 936 AGCCCGGAGGCGGCTCCAGCGGCGACTGGGGCTCTACTCTCCCAAGTTCCCAAGCT 995  
 Qy 180 GlySerLeuSerProAlaAlaSerLeuGIuArgProGIuLeuLeuGIyAlaThrSer 199  
 Db 996 GGTAGCTGAGCCCGACAGCTCTATTGGAGAGTTCCCTGGCTCGAGGTCCCAAGCTCC 1055  
 Qy 200 SerAlaCySleuSerProGIuSerLeuAlaPheSerAPpHeLeu 214  
 Db 1056 CCATCTGTCTGCTCCCGGAGACCTCGTGTCTCTGACTTCTTG 1100

RESULT 4  
 AAV27050  
 ID AAV27050 standard; cDNA; 804 BP.

AC AAV27050;  
 XX  
 DT 17-SEP-1998 (first entry)  
 XX  
 DE Mouse neurogenin 3 gene.  
 XX  
 KM ds; Mouse; neurogenin; expression vector; recombinant protein;  
 XX antibody; neurogenesis.  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 160..804  
 FT /tag= a  
 FT /product= "Mouse neurogenin 3"  
 XX  
 PN WO9813491-A2.  
 PD 02-APR-1998.  
 XX  
 PF 24-SEP-1997; 97WO-US17048.  
 XX  
 PR 17-SEP-1997; 97US-0932411.  
 PR 27-SEP-1996; 96US-0722570.  
 PR 12-NOV-1996; 96US-0030864.  
 PR 19-DEC-1996; 96US-0772009.  
 XX  
 PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 XX  
 PI Anderson DJ, Ma Q, Sommer L;  
 XX  
 DR MPI: 1998-230702/20.  
 DR P-PSDB; AAW54947.  
 XX  
 PT Mouse neurogenin, useful in neurogenesis - and recombinant nucleic  
 PT acids and proteins derived from rat and Xenopus  
 XX  
 PS Disclosure; Fig 9; 10pp; English.  
 XX  
 CC The Mouse neurogenin 3 is one of several neurogenin proteins discussed  
 CC in the present invention. The neurogenin nucleic acids can be expressed  
 CC in a host cell, transformed using an expression vector, to produce  
 CC recombinant proteins. The proteins and the antibodies raised against  
 CC the proteins are useful in the study of neurogenesis.  
 XX  
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

## Alignment Scores:

Pred. No.: 1,35e-42 Length: 804  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 19 Gaps: 2

US-09-595-947E-10 (1-214) x AAV27050 (1-804)

Qy 1 MetThrProGlnProSerGIyAlaProThrValGlnValThrArgGIuThrGIuArgSer 20  
 Db 160 ATGGGCGCTCATCCTTGGATGCGCTGCTCACTCAAGAGTCCCGAGAGCAACAACCT 219  
 Qy 21 PheProArgAlaSerGIuAspGIuValThrCysProThrSerAlaProProSerProThr 40  
 Db 220 TTTCGGAGGCTCGAGACACGAAGTGTCACTTCAATTCCACCCCACTTGGCCCACT 279  
 Qy 41 ArgThrProGIuAsCysAlaGluAlaGIuGIuGIyGIyCysArgGIyAlaProArgIyS 60  
 Db 280 CTCATACCTAGGAGACTGCTCCGAACAGAGTGGGTGACTGGCGAGGACTCGAGGAAG 339  
 Qy 61 LeuArgAlaArgArgGIyGIyArgSerArgProIySerGIuLeuAlaLeuSerIySgin 80  
 Db 340 CTCGGCGCCGACGCGGAGGCGGCAACAGGCCCAAGAGAGTGGCACTCGCAACACAG 399  
 Qy 81 ArgArgSerArgArgIyValAlaHisMetArgGIuArgArgMetHisAspLeuAsn 100  
 Db 400 CGAAGAGCGCGGCAAGAGGCCAATATGCGGAGCCCAATCGCATGCACAACTCAAC 459  
 Qy 101 SerAlaLeuAspAlaLeuArgGIyValLeuProThrPheProAspAlaIySleuThr 120  
 Db 460 TCGGCGCTGATGCGCTGCGCGGTCTCTGCCCACTTCCCGAGAGACCCAACTTACA 519  
 Qy 121 LysIleGIuThrLeuAArgPheAlaHisAntYrIleTPrAlaLeuThrGlnThrLeuArg 140  
 Db 520 AAGATCGAAGACCTGCGCTTCCGCCACACTACATCTGGGCACTGACAGCGTCCGC 579  
 Qy 141 IleAlaAPHisSerLeuTyrAlaLeuGIuProProlAProHicCySGlyGIuLeuGIy 160  
 Db 580 ATAGCGGACCAAGCTTCAACGCGCCGAGCCCTGTCCTCC---TGTGGAGAGCTGGG 636  
 Qy 161 SerPro---GlyGIuProProGIuAsPTrGIuSerLeuTyrSerProValSerGlnAla 179  
 Db 637 AGCCCGGAGGCTCCAGCGGAGACTGGGCTCTACTCTCCCAAGTCCCAAGCG 696  
 Qy 180 GlySerLeuSerProAlaAlaSerLeuGIuArgProGIuLeuLeuGIyAlaThrSer 199  
 Db 697 GGTAGCTGAGCCCGACAGCTCTATTGGAGAGTCCCTGGCTCGAGGTCCCAAGCTCC 756  
 Qy 200 SerAlaCySleuSerProGIuSerLeuAlaPheSerAPpHeLeu 214  
 Db 757 CCATCTGTCTGCTCCCGGAGACCTCGTGTCTCTGACTTCTTG 801

## RESULT 5

AAZ51981 standard; DNA; 804 BP.

ID AAZ51981;  
 AC AAZ51981;  
 XX  
 DT 04-JUL-2000 (first entry)  
 XX  
 DE Murine neurogenin-3 (NGN3) nucleic acid sequence.  
 XX  
 KM Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;  
 KM Phox2a protein; neuronal subtype-specific marker; growth factor;  
 KM neural differentiation; transplantation; neuronal dysfunction;  
 KM optical nerve damage; auditory nerve damage; neurodegenerative disorder;  
 KM neuroprotective; neurotropic; anticonvulsant; antiParkinsonian; vulnerary;  
 KM cerebroprotective; immunosuppressant; antiinfectious; ds.  
 XX  
 OS Mus sp.

Key Location/Qualifiers  
 CDS 160..804  
 /tag= a  
 /product= "Murine neurogenin-3 protein"  
 WO200009676-A2.  
 24-FEB-2000.  
 13-AUG-1999; 99WO-US18525.  
 14-AUG-1998; 98US-0096630.  
 (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 Anderson DJ, Lo L;  
 WPI; 2000-256250/22.  
 P-PSDB; AAY70570.  
 Inducing non-neuronal cells to differentiate into neurons and for  
 PT non-neuronal cells to express a neuronal subtype-specific marker,  
 PT comprising contacting the non-neuronal cells with a vector containing  
 PT neurogenin nucleic acid -  
 Claim 1; Fig 1J; 76pp; English.

The patent discloses a method for inducing non-neuronal cells (NNC) to  
 CC differentiate into neurons and for NNCs to express a neuronal subtype  
 CC -specific marker. Transformed host cells are used as sources of neuronal  
 CC and other growth factors; in culture for screening compounds that  
 CC modulate neural differentiation or as sources of recombinantly produced  
 CC neurogenins and phoxa proteins for use in transplantation. The cells  
 CC also have a variety of in vivo uses, e.g. for transplantation at sites of  
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to  
 CC optical or auditory nerve damage, brain or spinal cord injuries, and  
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present  
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor,  
 CC NNCs differentiate into neurons through the recombinant expression of a  
 CC transcription factor that induces a core program of neurogenesis. Forced  
 CC expression of murine NGN3 can elicit expression of at least some neuronal  
 CC phenotypic markers even in NNCs.

Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Alignment Scores:  
 Pred. No.: 1.35e-42 Length: 804  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 21 Gaps: 2

US-09-595-947E-10 (1-214) x AA251981 (1-804)

1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 160 ATGGGCTTCATCTTGGATGGCTCACCACCAAGTGTCCCGAGACACAACT 219  
 21 PheProArgAlaSerGlnuAspGluValThrCysProThrSerAlaProProSerProThr 40  
 220 TTTCCCGAGCTCGGACGACGAGAGTGTCCATTCACATTCACCCCACTGACCCACT 279  
 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgIys 60  
 280 CTCATACCTTGAAGGACTCTCCAGAGCAAGTGGGTGACTGCCGAGGAGCTTCGAGGAAG 339  
 61 LeuArgAlaArgArgGlyValArgSerArgProIysSerGlnLeuAlaLeuSerIysGln 80  
 340 CTCGGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 399  
 81 ArgArgSerArgArgIysValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

DB 400 CGAAGACCGGCGGAGAGGCGCAATGATCGGAGCGGCAATCGATGCACAACTCAAC 459  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaValLeuThr 120  
 DB 460 TCGGGCTGAGATGCGCTGCGGCTGCTCCGACCTTCGCGATGAGCGCAAACTTCA 519  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPAlaLeuThrGlnThrLeuArg 140  
 DB 520 AAGATCGAGACCTTCGCTGCTCCGACCAATCACTGCGGACATCACTGAGCGCTGCC 579  
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyValLeuGly 160  
 DB 580 ATAGCGGACCAAGCTTCTATAGCCCGGAGCCCTGTAGCCC---TGAGAGAGCTGGGG 636  
 QY 161 SerPro---GlyGlyProProGlyValAspThrProGlySerLeuTyrSerProValSerGlnAla 179  
 DB 637 AGCCCGGAGGTGCTCCACGCGGAGCTGAGCTTATCTACTCTCCAGTCTCCACAGCG 696  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyValLeuGlyValaThrSer 199  
 DB 697 GGTACCTGAGGCCCGACGCGCTCATTTGAGAGAAATTCCTGGCTGACAGTGCACGCTCC 756  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 757 CCATCTATCTGCTCCGAGAGACTGCTGTTCTCAGACTTCTTG 801  
 RESULT 6  
 AAF27266  
 ID AAF27266 standard; cDNA, 861 BP.  
 AC AAF27266;  
 DT 24-APR-2001 (first entry)  
 DE Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.  
 XX Atonal; homologue; orthologue; atonal-associated protein; deafness;  
 XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
 XX cellular proliferation; cerebellar granule neuron; gene therapy;  
 XX mechanoreceptive cell growth; auditory; osteopathic; cyostatic;  
 XX transgenic animal; ss.  
 OS Mus musculus.  
 XX WO200073764-A2.  
 XX 07-DEC-2000.  
 XX 01-JUN-2000; 2000WO-US15410.  
 XX 01-JUN-1999; 99US-0137060.  
 XX 19-JAN-2000; 2000US-0176993.  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 XX Zoghbi HY, Bellin H, Birmingham N, Hassan B, Ben-Arie N;  
 DR WPI; 2001-032190/04.  
 DR P-PSDB; AAB60359.  
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation -  
 PS Disclosure; Page -; 142pp; English.  
 CC The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologues or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal.

CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from Genbank.

XX Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

# Alignment Scores:

Pred. No.:	1,45e-42	Length:	861
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	22	Gaps:	2

US-09-595-947E-10 (1-214) x AAF27266 (1-861)

```

OY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGlnThrGlnArgSer 20
DB 160 ATGGGGCCCTCATCCCTTGATGCGCTCACCATCCAGTGTCCCGAGACACACACACT 219
OY 21 PheProArgAlaSerGlnAspGlnValThrCysProThrSerAlaProProSerProThr 40
DB 220 TTTCGCGAGCTCGGACGACGAGTGTCAATTCACCCACCTAGCCCACT 279
OY 41 ArgThrProGlnAsnGlyAlaGlnGluGlnGlyGlyCysArgGlyAlaProArgLys 60
DB 280 CTCATACCTAGGAGCTGCTCCGAGCAGAAAGTGGGCTGCGAGGAGCTCGAGGAAG 339
OY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGlnLeuAlaLeuSerLysGln 80
DB 340 CTCGCGCGCCGAGCGGAGGCGGACGAGCCGACGAGGAGTGGCACTCAGCAACAG 399
OY 81 ArgArgSerArgArgGlyLysAlaAsnAspArgGlnArgAsnArgMetHisAspLeuAsn 100
DB 400 CGAAGAGCGCGGCGGAGAGGCGCATGATCGGAGCGCATGCGATGCAACACTCAAC 459
OY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 460 TCGGGCGCTGATGCGTGGCGGTGCTGCGCCACCTTCGCGATGACCGCAAACTTACA 519
OY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGlnThrLeuArg 140
DB 520 AAGATGAGAACCTTGCGCTTCCGCCACATCATTTGGGCACTGACTGAGCGCTGCC 579
OY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 580 ATAGCGGAGCACAGACTTCTATGCGCCGCGAGCCCTCTGTGCC---TGTGAGAGGCTGGGG 636
OY 161 SerPro---GlyGlyProProGlnAspTyrPylSerLeuTyrSerProAlaSerGlnAla 179
DB 637 ACCCCCGGAGGTGGCTCCACAGCGGAGCTGGGGCTCTATCTACTCCCACTCTCCCAAGG 696
OY 180 GlySerLeuSerProAlaLaserLeuGlnGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 697 GGTAACTAGAGCCCAAGCGGCTCATTTGGAGAAATTCCTGGGCTGAGAGTCCCACTCC 756
OY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 757 CCATCTATCTGCTCCCGGAGGACACTGATGTTCTCAGACTTCTTG 801

```

RESULT 7  
 ABV75970  
 ID ABV75970 standard; cDNA; 861 BP.

XX ABV75970;  
 AC 11-FEB-2003 (first entry)  
 XX  
 DT Mouse transcription factor neurogenin 3 cDNA.  
 XX  
 DE Mouse; transcription factor; neurogenin 3; ngn3; stem cell;  
 XX differentiation; beta-cell; insulin; diabetes; hyperglycaemia;  
 KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;  
 KW gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 160..804  
 FT /tag=a  
 FT /product="Mouse ngn3"  
 FT  
 WO200286107-A2.  
 PD 31-OCT-2002.  
 XX  
 PF 19-APR-2002; 2002MO-EP04362.  
 XX  
 PR 19-APR-2001; 2001US-284531P.  
 XX  
 PA (DEVE-) DEVELOPENTWICKLINGSBIOLOGISCHE FORSCH.  
 PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.  
 XX  
 PI Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;  
 XX WPI; 2003-075629/07.  
 DR  
 XX  
 PT Differentiating stem cells into insulin-producing cells useful for  
 PT treating pancreatic diseases, by culturing stem cells in suitable  
 PT medium and activating gene involved in beta-cell differentiation -  
 XX  
 PS Disclosure; Page 58-59; 62pp; English.

CC The present sequence is that of cDNA encoding the murine  
 CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),  
 CC a gene which is required for the specification of the early  
 CC endocrine precursor in the pancreatic epithelium and which is  
 CC down-regulated once endocrine differentiation begins. The invention  
 CC provides a claimed method for differentiating stem cells (especially  
 CC embryonic, adult or somatic stem cells and primordial germ cells)  
 CC into insulin-producing cells. This involves culturing stem cells in  
 CC a suitable medium and activating at least one gene involved in  
 CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,  
 CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the  
 CC delivery of the gene into stem cells using a viral delivery  
 CC system, or the delivery of a protein product of the gene into stem  
 CC cells. The insulin-producing cells can be transplanted into  
 CC animals or human for treatment of pancreatic diseases, metabolic  
 CC syndrome and metabolic disorders with impaired glucose tolerance  
 CC as diabetes, hyperglycaemia and impaired glucose tolerance  
 CC (claimed). The cells can also be used to identify compounds which  
 CC stimulate beta-cell differentiation, insulin secretion or glucose  
 CC responsiveness. Differentiated beta-cells can be used to study the  
 CC toxic and other effects of exogenous compounds on beta-cell  
 CC function. In an example from the invention, Pax6 cDNA was inserted  
 CC into expression vector PACCMV, plpa under the control of the  
 CC cytomegalovirus promoter.

XX Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

# Alignment Scores:

Pred. No.:	1,45e-42	Length:	861
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2

DB:	25	Gaps:	2
US-09-595-947E-10 (1-214) x ABV75970 (1-861)			
QY	1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrArgSer 20		
Db	160 AAGGGGCTCATGCCCTTGATGGCTCACCATTCAGTGTGCCCCAGACACAAACCT 219		
QY	21 PheProAlaGlyAlaSerGluAlaGluValThrCysProThrSerAlaProProSerProThr 40		
Db	220 TTTCCCGGAGCTCCGAGCACGAAAGTGCAGTTCACATTCACACCACCTAGCCCACT 279		
QY	41 ArgThrProGlyAsnCysAlaGluValGluGluGluGlyCysArgGlyAlaProArgLys 60		
Db	280 CTCATATCTTAGGGAGCTGCTCCGAAAGAGAAAGTGGTGAATGCTCCGAGGGACCTGAGGAG 339		
QY	61 LeuArgAlaArgArgGlyGlyAlaArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80		
Db	340 CTCGGCGCCCGACGCCGAGGGGCCCAACAGGCCCAAGACGAGTGGCACTCAGCAACAG 399		
QY	81 ArgArgSerArgArgLysLysAlaAsnAspArgGluAlaGlnaGlnMetHisAspLeuAsn 100		
Db	400 CGAAGAGAGCCGGCGCAAGAGGCCATGATCGGAGGCGCAATGCATGACCAACTCAAC 459		
QY	101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120		
Db	460 TCGGCGGTGATGCGGTGGCGGGTGTCTGCCCACTTCCCGGATACAGCCCAACTTACA 519		
QY	121 LysIleGluThrLeuArgPheAlaHisAsnTyrlleThrAlaLeuThrGlnThrLeuArg 140		
Db	520 AAGATCGAAGCCCTGGGCTTCCGCCCAACTACATCTGGGCACTGACTGAGAGCTGGCG 579		
QY	141 IleAlaAspHisSerLeuTyrlaLeuGluProProAlaProHisCysGlyGluLeuGly 160		
Db	580 ATAGCGGACCAAGCTTATAGCCCGGAGGCCCTGTGTGCC---TGTGAGAGCTGGGG 636		
QY	161 SerPro---GlyGlyProProGlyAspTTPGlySerLeuTyserProValSerGlnAla 179		
Db	637 AGCCCCGAGGGTGGCTCCAAACGGGGACTGGGGCTCATCTACCTCCACAGCTCCCAAGCG 696		
QY	180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGluGlyAlaThrSer 199		
Db	697 GGTAACTAGACCCCAACGGCTCTCATTTGAGAGAAATTCCTGGGCGTGAAGTCCCAAGCTCC 756		
QY	200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214		
Db	757 CCATCTATCTGCTCCCGGAGACAGTGGTGTTCAGACTTCTTG 801		
RESULT 8			
ID	AAD46872		
XX	AAD46872 strand; DNA; 1860 BP.		
XX	AAD46872;		
XX	27-JAN-2003 (first entry)		
XX	DT		
XX	DE		
XX	Murine neurogenin 3 (Ngn3) gene.		
KW	Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;		
KW	type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;		
KW	islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; de.		
XX	Mus musculus.		
OS			
XX			
XX	Key	location/Qualifiers	
XX	FT	1093..1737	
XX	FT	CDS	
XX	FT	/tag= a	
XX	FT	/product= "Murine Ngn3 protein"	
XX	PN	WO200274045-A2.	
XX	PD	26-SEP-2002.	
XX			

```
PF      20-MAR-2002; 2002MO-US11166.
XX XX   20-MAR-2001; 2001US-0817360.
XX XX   (REGC ) UNIV CALIFORNIA.
XX PA
XX PI     German MS, Lin J;
XX DR     WPI; 2002-759853/82.
XX P-PSDB; AAE23278.
PT       Producing a mammalian islet cell for treating diabetes mellitus
PR    comprises introducing into a mammalian cell a nucleic acid molecule
PE    encoding neuroendocrine basic helix-loop-helix transcription factor -
XX XX
XX PS      Example 3; Page 89-90; 108bp; English.
CC CC     The invention relates to a method for producing a mammalian islet cell.
CC CC     The method comprising introducing into a mammalian cell a nucleic acid
CC CC     molecule encoding an islet transcription factor for expression of the
CC CC     islet transcription factor in the cell and for production of islet cell
CC CC     phenotype in the cell. The islet transcription factor is a neuroendocrine
CC CC     basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC CC     for treating type 2 diabetes mellitus and for replacing beta cells lost
CC CC     to autoimmune destruction in individuals with type 1 diabetes. The method
CC CC     is useful in cell therapy. The present sequence is murine neurogenin 3
CC CC     (Ngn3) gene.
SQ      Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;

Alignment Scores:
Pred. No.:          3,21e-42           Length:         1860
Score:              819.00             Matches:          163
Percent Similarity: 82.33%             Conservative:    14
Best Local Similarity: 75.81%           Mismatches:      36
Query Match:        72.67%             Indels:          2
Db:                 24                  Gaps:            2

US-09-595-947E-10 (1-214) x AAD46872 (1-1860)

QY      1 MetthrPrrogiInProserGlyAlaProthrValGluValThrArgGluThrGuArSer 20
Dd      1093 ATGGCGCTCATTCCTTGATGCCCTCACCAACAGTGTCCCAGAACAACAACACT 115
QY      21 PhetrProArGalASerGiUAspGlUValThrcySProThrsERAlapRProSerpProthr 40
Dd      1153 TTTCGCCGAGCCTCGAACAAGAAGTGTCAATTGCACCACCACCTAGCCCCACT 121
QY      41 ArgThrPrrogiYaenCyAAIagLuAlguLugLyGlyCyVarGilyAlaPbroArGlys 60
Dd      1213 CTCATVACTTAGGACTGTGCTCGAAGCAGAAATGGGTGACTCTCGAGGGACTTCGAAGAAG 127
QY      61 LeuArGalaaAgArGlyGlyVaRserArGrproLYsserglULeuAlaleusertysglin 80
Dd      1273 CTCCGCGCCACACCGGAGGGCGGAACAGGCCCAAAGACGATGGCACTCAGCAAAACG 133
QY      81 ArGaRtSeRaTgAtGLySLySAIaAsnAPsArRGluArGAAnArGMethIASpleuehn 100
Dd      1333 CGAAGAACCGCGCGCAAGAGAGCCCAATATATGGAGCCCAATTCGATCAACAACCTCAAC 139
QY      101 SerAlaleuaPaPalAleuarGlyValleubProthrPhetRoasPApaAlalyseuthr 120
Dd      1393 TCGGCGGTGGATGGTGCGTGGCGGTGTCTCTGCCACCTTCGCGATGAAGCCAAACTTA 145
QY      121 LyaIlegluThrLeuAlarGphelaAHIsAsntYrtLetRpalaLeuthrGlnThrlenuarg 140
Dd      1453 AAGATTCAGAGACCCCTGCGCTTGCACCACAATCAATCTGGGACTGACTCAACAGCTGGC 151
QY      141 IlealaspHisserLeutyRALaleuGIuPProPoaAlaPROHIscYEGLIGLUdeungly 160
Dd      1513 ATAAGCGAACAACAAGCTTTCTATGAGCCCGGAGGCCCTGTGGCC---TGTGGAGAGCTGGGG 156
QY      161 SerPro--GlyglyProProgiYAAPTpglyserLeutyRserProvalserGlnAla 179
```



```

Db      1570 AGCCCCGAGGTGGCTCCCAACGGGGAGCTGGGCTCTATCTACTCCCAAGTCCCAAGCG 1629
Qy      160 GYSSerLeuSerProAlaAlaSerLeuGluGluAArgProGlyLeuLeuGlyAlaThrSer 199
Db      1630 GGTAACTTGAGCCCAAGCGCTCATTTGGAGAAATTCCTCGGCTGAGGTGCCAGCTCC 1689
Qy      200 SerAlaCyLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      1690 CCATCTATCTGCTCTCCCGGAGACACTGGTGTCTCAGACTTCTTG 1734

RESULT 9
AAC61090
ID      AAC61090 standard; DNA; 1861 BP.
XX      AAC61090;
AC      AAC61090;
DT      05-FEB-2001 (first entry)
DE      Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX      Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
XX      islet cell precursor identification; mouse; ds.
OS      Mus musculus.
XX      Key      Location/Qualifiers
XX      CDS      1093..1737
XX      FT      /tag= a "Ngn3"
XX      FT      /product= "Ngn3"
XX      FT      /note= "Neurogenin 3"
XX      WO200059936-A1.
XX      12-OCT-2000.
XX      28-MAR-2000; 2000WO-US08436.
XX      06-APR-1999; 99US-0128180.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI; 2000-664989/64.
XX      P-PSDB; AAY85618.
XX      Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX      them, useful for diagnosis, prevention and treatment of diabetes
XX      mellitus and to identify individuals at risk of diabetes -
XX      Claim 18; Page 49-50; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX      protein AAY85617. The Ngn3 gene is located at chromosome position
XX      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX      protein sequences, and includes an antibody recognising the Ngn3 protein.
XX      Also included in the invention is a method for identifying an islet cell
XX      precursor, the method involves analysing a cell for the expression of the
XX      Ngn3 gene product, where detection of the product is indicative of an
XX      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX      reagent for detecting (in a subject) a predisposition to a defect in
XX      pancreatic islet cell function or formation associated with a defect in
XX      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX      precursor cells expressing Ngn3, and to alter cellular differentiation in
XX      culture in vivo to produce new beta-cells to treat patients with diabetes
XX      mellitus. The present sequence represents the murine Ngn3 genomic DNA
XX      sequence.
SQ      Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;
Alignment Scores:
Pred. No.: 3.21e-42 Length: 1861

```

```

Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mematches: 36
Query Match: 72.67% Indels: 2
DB: 21 Gaps: 2
US-09-595-947E-10 (1-214) x AAC61090 (1-1861)
Qy      1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db      1093 ATGGCGCTCATCCCTTGGATGCGCTCACCACTCAAGTGTCCCGAGACACAAACACT 1152
Qy      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db      1153 TTTCCCGAGCTCGGACCAAGAGTCTCAGATTCATTCACCCCACTAGCCCACT 1212
Qy      41 ArgThrProGlyAsnCyAlaGluAlaGluGluGlyGlyValArgGlyAlaProArgLys 60
Db      1213 CTCATACCTTAAGGACTGCTCCGAGCAGAGTGGTGACTCCGAGGAGCCTCGAGAG 1272
Qy      61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db      1273 CTCGCGCCCAAGCGGAGGCGGACAGGCCCAAGACGAGTGGCACTCAGCAAAAG 1332
Qy      81 ArgArgSerArgArgGlyValAlaAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db      1333 CGAAGAGCCGCGGCGAGAGAGCCCAATGATCGGAGCGCAATCGCATCACAACCTCAAC 1392
Qy      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Db      1393 TCGGCGCTGAGTGCCTCGCGGTCTCTGCGCCACTTCCGAGATGAGCCAAACTTACA 1452
Qy      121 LysIleGluThrLeuArgPheAlaHisAsnTrpIleThrAlaLeuThrGlnThrLeuArg 140
Db      1453 AAGATCGAGACCCCTGCGCTTGGCCCACTACATCACTGGCACTGACTCAGACGCTGGGC 1512
Qy      141 IleAlaAspHisSerLeuTrpAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db      1513 ATAGCGGACCAAGACTTATAGCCCGGAGCCCTCTGACC---TGTGAGAGACTGGGG 1569
Qy      161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTrpSerProValSerGlnAla 179
Db      1570 AGCCCCGAGGTGGCTCCCAACGGGAGCTGGGCTCTATCTACTCCCAAGTCCCAAGCG 1629
Qy      180 GYSSerLeuSerProAlaAlaSerLeuGluGluAArgProGlyLeuLeuGlyAlaThrSer 199
Db      1630 GGTAACTTGAGCCCAAGCGCTCATTTGGAGAAATTCCTCGGCTGAGGTGCCAGCTCC 1689
Qy      200 SerAlaCyLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      1690 CCATCTATCTGCTCTCCCGGAGACACTGGTGTCTCAGACTTCTTG 1734

RESULT 10
AAF27254
ID      AAF27254 standard; cDNA; 5567 BP.
XX      AAF27254;
AC      AAF27254;
DT      24-APR-2001 (first entry)
DE      Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.
XX      Atonal; homologue; orthologue; atonal-associated protein; deafness;
XX      hearing impairment; vestibular effect; balance disorder; osteoarthritis;
XX      cellular proliferation; cerebellar granule neuron; gene therapy;
XX      mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
XX      transgenic animal; ss.
OS      Mus musculus.
XX      WO200073764-A2.
XX      07-DEC-2000.

```



XX 01-JUN-2000; 2000WO-US15410.  
 XX 01-JUN-1999; 99US-0137060.  
 PR 19-JAN-2000; 2000US-0176993.  
 XX (BAYU) BAYLOR COLLEGE MEDICINE.  
 PA Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
 PI WPI: 2001-032190/04.  
 XX P-PSDB; AAB60350.  
 DR Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation -  
 XX Disclosure; Page -: 142pp; English.  
 PS The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologues or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal,  
 CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from Genbank.  
 XX SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;

Alignment Scores:  
 Pred. No.: 9, 91e-42 Length: 5567  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.338 Conservative: 14  
 Best Local Similarity: 75.818 Mismatches: 36  
 Query Match: 72.674 Indels: 2  
 DB: Gaps: 22

US-09-595-947E-10 (1-214) x AAF27254 (1-5567)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGlnArgSer 20  
 DB 4923 ATGGCGCCCTCATCCCTTGATGGCTCACCATCCAGATGTCGCCAGACACAACT 4982  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 4983 TTTCCTCGAGCTTCGAGCAGCAAGTGTCACTTCATTCCACCCCACTTGGCCCACT 5042  
 QY 41 ArgThrProGlyAsnGlyAsnGluGluGluGluGluGluGluGluGluGluGluGluGlu 60  
 DB 5043 CTCATACCTAGGAGCTCTCCAGAGCAAGAGTGGTGAATGCTCCGAGGAGCTCGAGGAG 5102  
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerArgGln 80  
 DB 5103 CTCCTCGCCCGGAGCGCGAGGCGGCAAGGCCCAAGAGCGAGTGGCACTCAGCAAAACG 5162  
 QY 81 ArgArgSerArgArgGlyGlyGlyAlaLeuAspArgGluArgAspArgMetHisAspLeuAn 100  
 DB 5163 CGAAGAGCGCGGCGAAGAGCGCAATGATCGGAGCGCAATCGCATGACACCACTCAAC 5222  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120

DB 5223 TCGGCGCTGATGCCCTCGGCGGTCTCTGCTCCGACCTTCGCGAGTACGCCAACTTACA 5282  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTrpIleTPAlaLeuThrGlnThrLeuArg 140  
 DB 5283 AAGATCGAGACCTCGCTGCTGCGCCCACTACATCTGGACCTAGCGCTGCGCC 5342  
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 DB 5343 ATAGCGGACCAAGCTTATATGCGCCGAGGCCCTGTGCCC---TGTGGAAGCTGGGG 5399  
 QY 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuThrYserProValSerGlnAla 179  
 DB 5400 AGCCCGGAGGTGGCTCCCAAGGAGACTGGGGCTTATCTACTCCCAAGTCCCAAGCG 5459  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGluAlaThrSer 199  
 DB 5460 GGTAACTGAGCCCAAGCCCTCATTTGAGAGGATTTCCCTGGCTTACAGGTGCCAGCTCC 5519  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 5520 CCATCTATCTGCTCCCGGAGACCTGGTGTCTCAGACTTCTTG 5564

RESULT 11  
 ABQ49522/C  
 ID ABQ49522 standard; DNA; 592 BP.  
 XX ABQ49522;  
 XX 12-JUL-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.  
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KW drug; side effect; cancer; central nervous system; cardiovascular;  
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KW SNP; cell differentiation; ds.  
 XX Homo sapiens.  
 OS WO200218632-A2.  
 XX PD 07-MAR-2002.  
 XX PD 01-SEP-2001; 2001MO-BP10074.  
 XX PR 01-SEP-2000; 2000DE-1043826.  
 XX PR 05-SEP-2000; 2000DE-1044543.  
 XX PA (EPIC-) EPIGENOMICS AG.  
 XX PI Olek A, Piepenbrock C, Berlin K, Guettig D;  
 DR WPI: 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 PS Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of  
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 genomic sample of DNA. The sample is treated chemically to convert  
 cytosine (C) but not methylated C, to uracil, then part of the genomic  
 DNA that contains the target C is amplified to form a labeled amplicon.  
 The amplicon is hybridized to two classes, each with at least one  
 member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 and the degree of hybridisation to both classes is determined from the  
 label on the amplicon. From the ratio of labels hybridised to the two  
 classes of oligomers, the degree of methylation is calculated. The method  
 is used: (1) for diagnosis and/or prognosis of side effects of  
 therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

# Alignment Scores:

Pred. No.:	4,21e-24	Length:	592
Score:	519.00	Matches:	98
Percent Similarity:	83.67%	Conservative:	25
Best Local Similarity:	66.67%	Mismatches:	24
Query Match:	46.05%	Indels:	0
DB:	24	Gaps:	0

US-09-595-947E-10 (1-214) x ABQ49522 (1-592)

QY 1 MetTnProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db 443 ATAAAGCCTCAACCTCGAATACGCCCTATCCAAATACCCGTAACGAAACGATCC 384  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 Db 383 TTCCCCAAACCTCGAATAACGAAATACCTACCCACGTCGCCGCCGCAACCCCACT 324  
 QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60  
 Db 323 CGCACACGAAAAAATCTACGCAAAAAAAGAAAAAATACTACCGAAAAAATCCGAAAAA 264  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 263 CTCGCAACGACGCGCAAAAAAGCAACCACTTAAACGATTAACACTTAACAAACA 204  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 Db 203 CGACGAAATCGACGAAAAAACAACCAACGACCGCAACGCAATGCAATACCAACCTCAAC 144  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 Db 143 TCGACACTTAACGCTTACGCAATATCCATCCACTTCCCAACGACGCAAACTCACCC 84  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGlnThrLeuArg 140  
 Db 83 AAAATCGAAACGCTACGCTTCCGCCCAACACTACATCTTAAACGTAACCTAAACGCTACCC 24  
 QY 141 IleAlaAspHisSerLeuTyr 147  
 Db 23 ATTAACGACCAACCACTTATAC 3  
 RESULT 12  
 ABQ49523  
 ID ABQ49523 standard; DNA; 592 BP.  
 XX ABQ49523;  
 DT 12-JUL-2002 (first entry)  
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.  
 DE Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 XX drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 XX SNP; cell differentiation; ds.  
 OS Homo sapiens.  
 OS MO200218632-A2.  
 PN 07-MAR-2002.  
 PD

XX 01-SEP-2001; 2001WO-EPI0074.  
 XX 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX (EPIG-) EPIGENOMICS AG.  
 XX Olek A, Piepenbrock C, Berlin K, Guetig D,  
 PI WPI; 2002-371829/40.  
 DR

PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA  
 PS Claim 12; 56bp + Sequence Listing; 56bp; German.

CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridised to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridisation to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridised to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's) and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Alignment Scores:	4,21e-24	Length:	592
Pred. No.:	519.00	Matches:	98
Score:	83.67%	Conservative:	25
Percent Similarity:	66.67%	Mismatches:	24
Best Local Similarity:	46.05%	Indels:	0
Query Match:	24	Gaps:	0

US-09-595-947E-10 (1-214) x ABQ49523 (1-592)

QY 1 MetTnProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
 Db 150 ATAAAGCCTCAACCTCGAATACGCCCTATCCAAATACCCGTAACGAAACGATCC 209  
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40  
 Db 210 TTCCCCAAACCTCGAATAACGAAATACCTACCCACGTCGCCGCCGCAACCCCACT 269  
 QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60  
 Db 270 CGCACACGAAAAAATCTACGCAAAAAAAGAAAAAATACTACCGAAAAAATCCGAAAAA 329  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 Db 330 CTCGCAACGACGCGCAAAAAAGCAACCACTTAAACGATTAACACTTAACAAACA 389  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100  
 Db 390 CGACGAAATCGACGAAAAAACAACCAACGACCGCAACGCAATGCAATACCAACCTCAAC 449  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

DB 450 TCGACACTAAGCCCTACGCGATTCCTACCCACCTTCCCAAGACGCGAAGCTACCC 509  
 QY 121 LyeillegluthrLeuargPhealIahIsenTyrlIeTPalaleuThglrthleuarg 140  
 DB 510 AAAATCGAAACCGCTTCGCGCCACACTACCTTAACCGCTAACCGCTACGCG 569  
 QY 141 IleaIaaphIaSerleuTy 147  
 DB 570 ATTAACGAAACCACTTATAC 590  
 RESULT 13  
 ABO49524  
 ID ABO49524 standard; DNA; 592 BP.  
 AC ABO49524;  
 XX  
 DT 12-UTU-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.  
 XX  
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful  
 PT for diagnosis and prognosis, comprises selective hybridization of  
 PT amplicons from chemically treated DNA -  
 XX  
 PS Claim 12; 56pp + Sequence listing; 56pp; German.  
 XX  
 CC This invention describes a novel method for determining the degree of  
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 CC genomic sample of DNA. The sample is treated chemically to convert in a  
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
 CC DNA that contains the target C is amplified to form a labeled amplicon.  
 CC The amplicon is hybridized to two classes, each with at least one  
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 CC and the degree of hybridization to both classes is determined from the  
 CC label on the amplicon. From the ratio of labels hybridized to the two  
 CC classes of oligomers, the degree of methylation is calculated. The method  
 CC is used: (i) for diagnosis and/or prognosis of side effects of  
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
 CC systems etc., particularly by detecting mutations or single nucleotide  
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 CC types and for investigating cell differentiation. The method allows the  
 CC methylation status of many C residues to be determined simultaneously.  
 CC ABO13410-ABO54121 represent genomic DNA sequences used to illustrate the  
 CC method for determining the degree of cytosine methylation described in  
 CC the disclosure of the invention.  
 CC  
 SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 7.13e-21 Length: 592

Score: 467.00 Matches: 101  
 Percent Similarity: 73.94% Conservative: 4  
 Best Local Similarity: 71.13% Mismatches: 37  
 Query Match: 41.44% Indels: 0  
 DB: 24 Gaps: 0  
 US-09-595-947e-10 (1-214) x ABO49524 (1-592)  
 QY 6 SerGIYAlaProthrValGlnValThrArgGluArgSerPheProArgAlaSer 25  
 DB 165 TCGGTCGCTTATTTGTTTAAGTATTCGTGAGACGAGCGGTTTTTTTGAAGTTTCG 224  
 QY 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45  
 DB 225 GAAGACGAAGGATTTGTTTACGTTGCTTCGTTAGTTTATTCGTATACCGGGGAT 284  
 QY 46 CysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgGlyLeuArgAlaArg 65  
 DB 285 TGCCTAGAGGCGGAAGAGGAGGTTGTCAGAGGCGTTTCAGAGGAAGTTTCGGGTACGGCGC 344  
 QY 66 GlyGlyArgSerArgProGlySerGluLeuAlaLeuSerLysGlnArgArgSerArg 85  
 DB 345 GGGGACGCTAGTCGCTTTAAGACGAGTTGATGAGTAAAGTACGACGAGGTCGGCA 404  
 QY 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105  
 DB 405 AAGAAAGTTAAACGATCGGACGCTATCGAATGTATATTTTAAATTCGATTTGACGTT 464  
 QY 106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125  
 DB 465 TTGCGCGGTGTTGTTGTTATTTTATTTAGACACGCGAAGTTTATTAAGATCGAGCGTTG 524  
 QY 126 ArgPheAlaHisAsnTyrlIeTPalaleuThglrthleuargIleAlaAspHisSer 145  
 DB 525 CGTTTCGTTTAATTAATTAATTTGCGGCTTGATTTAAACGTTGCGTATGCGGATTAAGT 584  
 QY 146 LeuTy 147  
 DB 585 TTGTAC 590  
 RESULT 14  
 ABO49525/C  
 ID ABO49525 standard; DNA; 592 BP.  
 AC ABO49525;  
 XX  
 DT 12-UTU-2002 (first entry)  
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.  
 XX  
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 KM drug; side effect; cancer; central nervous system; cardiovascular;  
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;  
 KM SNP; cell differentiation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200218632-A2.  
 XX  
 PD 07-MAR-2002.  
 XX  
 PF 01-SEP-2001; 2001WO-EP10074.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 PR 05-SEP-2000; 2000DE-1044543.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;  
 DR WPI; 2002-371829/40.  
 PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA  
XX  
XX  
PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridised to two classes, each with at least one  
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridisation to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridised to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

XX  
XX Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

# Alignment Scores:

Pred. No.:	7,13e-21	Length:	592
Score:	467.00	Matches:	101
Percent Similarity:	73.94%	Conservative:	4
Best Local Similarity:	71.13%	Mismatches:	37
Query Match:	41.44%	Indels:	0
DB:	24	Gaps:	0

US-09-595-947E-10 (1-214) x ABQ49525 (1-592)

QY 6 SerrgiyalaProthValGlnValTharGluThrGluArgSerPheProArgAlaSer 25  
DB 428 TCGGGGCGCTTATTGTTTAAGTGAATTCGTAGACGACGCGTTTATTTTGAAGTTTCG 369  
QY 26 GluAerGluValThrCySerProthSerAlaProPheSerProthArgThrProGlyAn 45  
DB 368 GAAGACGAAGTATTGTTTACGTTTCGTTTACGTTTATTCGTTTACGCGGGGAAAT 309  
QY 46 CysAlaGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 65  
DB 308 TCCGTAAGAGCGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 249  
QY 66 GlyGlyArgSerArgProLysSerGluLeuAlaLeuSerGluAlaArgSerArgArg 85  
DB 248 GGGGGGCGTAGTCGGTTTAAAGCGAGTGTATTAAGTAAGTAAGTAAGTAAGTAAGTAAG 189  
QY 86 LysLysAlaAspAspArgGluArgGluArgGluArgGluArgGluArgGluArgGluArg 105  
DB 188 AAGAAAGTTAAGCATCGACGCGTATCAATGTAATTAATTAATTAATTAATTAATTAATTA 129  
QY 106 LeuArgGlyValLeuProthPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125  
DB 128 TTGCGCGGCTTTGTTTATTTTATTTTAAAGCGACGAGTTTATTAAGATTCAGACGTTG 69  
QY 126 ArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145  
DB 68 CGTTTCGTTTAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9  
QY 146 LeuTyr 147  
DB 8 TTGTAC 3

RESULT 15  
AAF27264  
ID AAF27264 standard; cDNA; 790 BP.

XX  
AC AAF27264;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Chicken atonal homologue ngn2/ath4 cDNA, SEQ ID NO:20.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;  
XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
XX cellular proliferation; cerebellar granule neuron; gene therapy;  
XX mechanoreceptive cell growth; auditory; osteopathic; cyostatic;  
XX transgenic animal; ss.

XX Gallus gallus.

XX WO200073764-A2.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.

XX 01-JUN-1999; 99US-0137060.

XX 19-JAN-2000; 2000US-0176993.

XX (BATU ) BAYLOR COLLEGE MEDICINE.

XX Zoghbi HY, Belien H, Birmingham N, Hassan B, Ben-Arie N;

XX WPI; 2001-032190/04.

XX P-PSDB; AAB60357.

XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
XX any of its homologs or orthologs, for the treatment of e.g. deafness,  
XX osteoarthritis and abnormal cell proliferation -

XX Disclosure; Page -; 142pp; English.

XX The invention relates to the use of atonal-associated nucleic acid or  
XX amino acid sequence, or any of its homologues or orthologues as  
XX therapeutic agents for the treatment of deafness, partial hearing loss,  
XX vestibular effects due to damage or loss of inner hair cells,  
XX osteoarthritis and abnormal cell proliferation. The invention also  
XX encompasses methods of screening for compounds which affect the  
XX expression of an atonal-associated nucleic acid sequence in an animal,  
XX and a transgenic animal in which an allele of a native atonal-associated  
XX gene is replaced by a heterologous nucleic acid sequence, thus  
XX inactivating the atonal-associated allele. The nucleic acids or proteins  
XX may be used in a method of treating an animal for hearing impairment,  
XX joint disease, balance disorders, abnormal cell proliferation, or other  
XX disease related to loss of a functional atonal-associated nucleic acid or  
XX protein. They may particularly be used to treat an animal with a  
XX deficiency in cerebellar granule neurons or their precursors, and may  
XX also be used in promoting mechanoreceptive cell growth and generating  
XX hair cells. The present sequence represents an atonal-associated nucleic  
XX acid sequence referred to in the invention.  
XX Note: The present sequence is not shown in the specification, but  
XX was obtained from GenBank.

XX Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

# Alignment Scores:

Pred. No.:	7,19e-16	Length:	790
Score:	388.50	Matches:	104
Percent Similarity:	54.31%	Conservative:	22
Best Local Similarity:	44.83%	Mismatches:	63
Query Match:	34.47%	Indels:	43
DB:	22	Gaps:	10

US-09-595-947E-10 (1-214) x AAF27264 (1-790)

QY 3 ProGlnProSerGlyAlaProthValGlnValTharGluThrGluArgSerPhePro 22  
DB 47 CCAAGCGCCCGCGACCGCGCGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 103



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 2, 2004, 18:55:26 ; Search time 355 Seconds  
(without alignments)  
2197.000 Million cell updates/sec

Title: US-09-595-947E-10

Perfect score: 1127  
Sequence: 1 MTPGSGAPTVGVTRETERS.....LGATSSACLSPEGLAFSDPL 214

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 243493 segs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09595947/runat\_02022004\_154934\_8079/app\_query.fasta\_1.391  
-DB=Published Applications NA -QWMT=fastcap -SUFFIX=trpb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=biosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USR=US09595947@cgn\_1\_1.387@runat\_02022004\_154934\_8079  
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:\*

1:	/cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2:	/cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3:	/cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4:	/cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5:	/cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6:	/cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7:	/cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8:	/cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9:	/cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
11:	/cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13:	/cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
14:	/cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15:	/cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16:	/cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17:	/cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18:	/cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
------------	-------	-------------	-----------	----	-------------

1	1105	98.0	5340	9	US-09-817-360-1	Sequence 1, Appli
2	819	72.7	645	14	US-10-004-717-4	Sequence 4, Appli
3	819	72.7	861	14	US-10-004-717-24	Sequence 24, Appli
4	819	72.7	1861	9	US-09-817-360-3	Sequence 3, Appli
5	388.5	34.5	790	14	US-10-004-717-20	Sequence 20, Appli
6	388.5	34.2	1074	14	US-10-004-717-18	Sequence 18, Appli
7	375.5	33.3	1385	14	US-10-004-717-30	Sequence 30, Appli
8	370.5	32.9	1412	14	US-10-004-717-6	Sequence 6, Appli
9	370.5	32.9	1412	14	US-10-004-717-37	Sequence 37, Appli
10	370	32.8	1527	8	US-08-722-570-12	Sequence 12, Appli
11	359.5	31.9	738	8	US-08-722-570-13	Sequence 13, Appli
12	322.5	28.6	1312	8	US-08-722-570-14	Sequence 14, Appli
13	304	27.0	1277	8	US-08-722-570-15	Sequence 15, Appli
14	238.5	21.2	1550	14	US-10-004-717-13	Sequence 43, Appli
15	238.5	21.2	1957	14	US-10-004-717-8	Sequence 8, Appli
16	215	19.1	1099	13	US-10-413-358-27	Sequence 27, Appli
17	215	19.1	1211	13	US-10-413-358-26	Sequence 26, Appli
18	207.5	18.4	993	14	US-10-004-717-47	Sequence 47, Appli
19	207.5	18.4	3261	14	US-10-004-717-12	Sequence 12, Appli
20	207.5	18.4	3541	14	US-10-004-717-32	Sequence 32, Appli
21	206	18.3	1021	12	US-10-321-039-71	Sequence 71, Appli
22	205.5	18.2	1056	14	US-10-004-717-10	Sequence 10, Appli
23	205.5	18.2	1393	14	US-10-004-717-45	Sequence 45, Appli
24	198.5	17.6	675	14	US-10-004-717-15	Sequence 15, Appli
25	198.5	17.6	2196	12	US-10-108-260A-1239	Sequence 1239, Ap
26	197.5	17.5	485	14	US-10-029-386-25014	Sequence 59, Appli
27	194	17.2	748	13	US-10-029-386-25014	Sequence 25014, A
28	192.5	17.1	1065	14	US-10-004-717-57	Sequence 1, Appli
29	192.5	17.1	1572	14	US-10-004-717-57	Sequence 57, Appli
30	191.5	17.0	2315	12	US-10-136-728-77	Sequence 77, Appli
31	190	16.9	501	14	US-10-004-717-13	Sequence 13, Appli
32	187	16.6	948	14	US-10-004-717-41	Sequence 41, Appli
33	187	16.6	2993	15	US-10-125-237-56	Sequence 56, Appli
34	187	16.6	2993	15	US-10-105-891-56	Sequence 56, Appli
35	185.5	16.5	849	13	US-10-226-872-8	Sequence 8, Appli
36	183.5	16.3	937	14	US-10-004-717-65	Sequence 65, Appli
37	183.5	16.3	902	12	US-10-004-717-1190	Sequence 1190, Ap
38	183.5	16.3	1791	13	US-10-226-872-1	Sequence 1, Appli
39	183.5	16.3	1791	13	US-10-226-872-6	Sequence 6, Appli
40	183.5	16.3	1830	13	US-10-226-872-4	Sequence 4, Appli
41	183.5	16.3	17290	11	US-09-999-121-7	Sequence 7, Appli
42	183.5	16.3	25760	11	US-09-999-121-13	Sequence 13, Appli
43	180	16.0	1034	13	US-10-029-386-22762	Sequence 22762, A
44	179	15.9	849	13	US-10-226-872-9	Sequence 9, Appli
45	177.5	15.7	938	14	US-10-004-717-39	Sequence 39, Appli

#### ALIGNMENTS

RESULT 1  
US-09-817-360-1  
Sequence 1, Application US/09817360  
Patent No. US20020015696A1  
GENERAL INFORMATION:  
APPLICANT: Lin, Joseph  
APPLICANT: German, Michael S.  
TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
TITLE OF INVENTION: AND DELIVERY OF INSULIN  
FILE REFERENCE: UCSF-129C1P  
CURRENT APPLICATION NUMBER: US/09/817.360  
CURRENT FILING DATE: 2001-03-20  
PRIOR APPLICATION NUMBER: 09/535,145  
PRIOR FILING DATE: 2000-03-24  
PRIOR APPLICATION NUMBER: 60/128,180  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 5340  
TYPE: DNA  
ORGANISM: Homo Sapiens  
US-09-817-360-1

## Alignment Scores:

Pred. No.: 6,88e-92 Length: 5340  
 Score: 1105.00 Matches: 211  
 Percent Similarity: 99.07% Conservative: 1  
 Best Local Similarity: 98.60% Mismatches: 2  
 Query Match: 98.05% Indels: 0  
 DB: 9 Gaps: 0

US-09-595-947E-10 (1-214) x US-09-817-360-1 (1-5340)

QY 1 MetThrProGlnProSerGlyValAProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 3022 ATGAGCGCTCAACCTCGGGTGGCCACTGTCACAGTGAACCGGTGAGACGAGCGGTCC 3081  
 QY 21 PheProArgAlaSerGluaspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 3082 TTCCCCAGAGCCTCGGAGACGAGTGAAGTCCGACGCTCGCCCGCCGCGCCGACT 3141  
 QY 41 ArgThrProGlyValAsnGlyAlaGluGluGluGlyCysArgGlyValAProArgLys 60  
 DB 3142 CCGCACCGGGGAACTGCCAGAGGCGGAGAGAGGCTGCCGAGGGGCCCGAGAGAG 3201  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 3202 CTCGGGCGACGGCGCGGGGAGACGACCGGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAProArgGluArgAsnArgMetHisAspLeuAsn 100  
 DB 3262 CCACGGAGTCGGCGGAGAAAGGCCACACCGGACCGGACATCGAATGCACACCTCAAC 3321  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 DB 3322 TCGGCACTGAGAGCGCTGCGCGGTGCTCTGCCACCTTCCAGACGACCGAAGCTCAAC 3381  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIleTrpAlaLeuThrGlnThrLeuArg 140  
 DB 3382 AAGATCGAAGACGCTGGCTTCCGCCACACTACATCTGGGCGCTGACTCAACGCTGCCG 3441  
 QY 141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisArgLysGluLeuGly 160  
 DB 3442 ATAGCGGACACACGCTTGTACGCGCTGAGCGCGCCGCGCACCTGCGGAGGCTGGGC 3501  
 QY 161 SerProGlyLysProProGlyLysPTrpGlySerLeuTyrlSerProValSerGlnAlaGly 180  
 DB 3502 AGCCACGAGGCTTCCCGCGGAGACTGGGGGTCCCTTACTCCCACTCTCCAGGCTGGC 3561  
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyValThrSerSer 200  
 DB 3562 ACCCTGAGTCCCGCCGCTGCTGAGGAGCGACCCGCGCTGCTGGGGGCCACTCTTCC 3621  
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 3622 GCGCTGTAGCCCGACGAGCTGGCTTCTCAGATTCTTG 3663  
 RESULT 2  
 US-10-004-717-4  
 ; Sequence 4, Application US/10004717  
 ; Publication No. US20020192665A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZOGBI, HUDA Y.  
 ; APPLICANT: YANG, QI  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
 ; FILE REFERENCE: P01899US4  
 ; CURRENT APPLICATION NUMBER: US/10/004,717  
 ; PRIOR FILING DATE: 2002-08-16  
 ; PRIOR APPLICATION NUMBER: 09/585,645  
 ; PRIOR FILING DATE: 2000-06-01  
 ; PRIOR APPLICATION NUMBER: 60/176,993  
 ; PRIOR FILING DATE: 2000-01-19  
 ; PRIOR APPLICATION NUMBER: 60/137,060  
 ; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO: 4  
 ; LENGTH: 645  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 US-10-004-717-4

## Alignment Scores:

Pred. No.: 1.4e-66 Length: 645  
 Score: 819.00 Matches: 163  
 Percent Similarity: 82.33% Conservative: 14  
 Best Local Similarity: 75.81% Mismatches: 36  
 Query Match: 72.67% Indels: 2  
 DB: 14 Gaps: 2

US-09-595-947E-10 (1-214) x US-10-004-717-4 (1-645)

QY 1 MetThrProGlnProSerGlyValAProThrValGlnValThrArgGluThrGluArgSer 20  
 DB 1 ATGGGCGCTCACTCCTTGATGCGCTCAACATCCAAAGTGTCCCGAGACACACACACT 60  
 QY 21 PheProArgAlaSerGluaspGluValThrCysProThrSerAlaProProSerProThr 40  
 DB 61 TTCCCCAGAGCCTCGAGACACGAGTGAAGTCCAGTTCCAAATCCACCCCACTGAGCCCACT 120  
 QY 41 ArgThrProGlyValAsnGlyAlaGluGluGlyCysArgGlyValAProArgLys 60  
 DB 121 CTCATACCTTACGAGACTGCTCCAGACGAGAGTGTGACTGCGAGGAGCTCGAGAGAG 180  
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80  
 DB 181 CTCGCCGCCGACGCGGAGGCGGACACAGGCCCAAGAGAGTGGCACTGCAACGAAACG 240  
 QY 81 ArgArgSerArgArgLysLysAlaAsnAProArgGluArgAsnArgMetHisAspLeuAsn 100  
 DB 241 CCAAGAGCCGCGCGCAAGAGGCCAATATCGGAGGCCCAATCGCATGACACAACTCAAC 300  
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120  
 DB 301 TCGGCGTGAAGCGCTGCGCGGTGCTCTGCCACCTTCCGAGTGAAGCGCAAACTTACA 360  
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIleTrpAlaLeuThrGlnThrLeuArg 140  
 DB 361 AAGATCGAAGACCTTGGCTTGGCCCAACTACATCTGGGACCTGACTCAGAGCTGGGC 420  
 QY 141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160  
 DB 421 ATAGCGGACACACGCTTATAGCCCGGAGCCCTGTGCCC--TGTGAGAGAGCTGGGG 477  
 QY 161 SerPro---GlyGlyProProGlyLysPTrpGlySerLeuTyrlSerProValSerGlnAla 179  
 DB 478 AGCCCCGAGGTGGCTCCAGCGGAGACTGGGGCTTACTTACTCCCACTTCCCAAGCG 537  
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyValThrSer 199  
 DB 538 GGTAACTGAGCCCAAGCGCTCATTTGAGAGAAATCTCCGTGGCTCAGAGTGGCCAGCTCC 597  
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214  
 DB 598 CATCTCATCTCTCCCGGAGACACTGGTGTCTCAGACTTCTTG 642  
 RESULT 3  
 US-10-004-717-24  
 ; Sequence 24, Application US/10004717  
 ; Publication No. US20020192665A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ZOGBI, HUDA Y.  
 ; APPLICANT: YANG, QI  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
 ; FILE REFERENCE: P01899US4





```

Db      1690 CCATCTATGCTCCCGGAGACACTGATGTTCTCAGACTCTTG 1734
RESULT 5
US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, OI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-20

Alignment Scores:
Pred. No.:      6,44e-27      Length:      790
Score:          388.50      Matches:     104
Percent Similarity: 54.31%      Conservative: 22
Best Local Similarity: 44.83%      Mismatches: 63
Query Match:    34,478      Indels:      43
DB:             14          Gaps:          10

US-09-595-947E-10 (1-214) x US-10-004-717-20 (1-790)
QY      3 ProGlnProSerGlyAlaProThrValGlnValThrArgGlnThrGluArgSerPro 22
Db      47 CCAGCGCGCCCGCCGAGCCCGCCGAGCGCC--GTCGACCAAGCGCGCCCGCCCTCTCC 103
QY      23 Arg-----AlaSerGlnuSpGluVal 29
Db      104 CCGCCCGGAGAGATCGCGGTGAAGGCGAGAGCCCGCGCGCGCGCGAGACCACTG 163
QY      30 ThrCysProThrSerAlaProProSerProThrArgThr--ProGlyAsnCysAlaGlu 48
Db      164 CTGCTGCTCGCTCGCTCGCTCGCGCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 223
QY      49 AlaGlnGlnGlyGlyCysArgGlyAlaProArgGlyLeuArg-----AlaArgArgGly 66
Db      224 GAGAGCGAGAGAGAGAGAGAGAGCGCGCGCGCGCGCTCGAGAGAGCGCTCGCGCGCG 283
QY      67 GlyArgSerArgProLysSerGlnuLeuAlaLeuSerGln-----Arg 81
Db      284 GGGCGGCGAGAGAGCGCGCGCGCGCGCGCGCGAGCGAGAGCGCGAGCGCATCAG 343
QY      82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
Db      344 CGGAGCGCGCGCTGAAGCGCAACACCGGAGCGGACCGCATGCAACCTGAAACGCG 403
QY      102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
Db      404 GCGCTGGAAGCGCTGGCGCAAGCGCTGCCACCTTCCCGAGGAGCCCAAGCTCAACGA 463
QY      122 TleGluThrLeuArgPheAlaHisAsnYrLietrAlaLeuThrGlnThrLeuArgGlie 141
Db      464 ATCGAAGCGCTGCGCTTGGCCCAACATTCATTCGCGCTCAACGAGAGCGCTGCGCTG 523
QY      142 AlaAspHisSer--LeuYrAlaLeuGlnuProProAlaProHisCysGlyGlnuLeuGly 160

```

```

Db      524 GCCCGGCGCGCGCTCGCGGAGCGCGCCGAGCGCGCGCGCGCGCGCGAGGCG 583
QY      161 SerProGlyGlyProProGlyAspThrGlySerLeuYrSerProValSerGlnuLagly 180
Db      584 AGCCCC--TCGCGCGCTCTGCTGCG-----AGCGGC 613
QY      181 SerLeuSerProAlaAlaSerLeuGlnuLagArgProGlyLeuLeuGlyAlaThrSerSer 200
Db      614 GCGCGCACCGCGCGCGCGCTCC-----GCTCGCGCTTAC 646
QY      201 AlaCys--LeuSerProGlySerLeuAlaPheSer 211
Db      647 GCTGCACTTATGCTCCGCGAGCGCGCGCGCGCTCC 682

RESULT 6
US-10-004-717-18
; Sequence 18, Application US/10004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, OI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-18

Alignment Scores:
Pred. No.:      1.69e-26      Length:      1074
Score:          385.50      Matches:     104
Percent Similarity: 51.29%      Conservative: 15
Best Local Similarity: 44.83%      Mismatches: 60
Query Match:    34,218      Indels:      53
DB:             14          Gaps:          10

US-09-595-947E-10 (1-214) x US-10-004-717-18 (1-1074)
QY      21 PheProAlaAlaSerGlnuSpGlnuValThrCysProThrSerAlaProProSerPro--- 39
Db      228 TTTCTCTCTCTTCTCTCTCTCTC-----CTCAGACACTTCTCTCTCTCTCTCTCT 275
QY      40 -----ThrArgThrProGlyAsnCysAlaGluAla 49
Db      276 GCCCGCCACTGACGCGCTCGGCTTCTCTCGGAGAGAGCGCGAGCGCGAGCGAGC 335
QY      50 GlnuGlnuGlyGlyCysArg-----GlyAlaProArgGlySerLeuArgAlaArgArgGly 66
Db      336 -----GGCGGCGTTTCGAGAGCGCGCGCGAGGCTCGCGGAG-----CGCGGAGAGAGCGC 386
QY      67 GlyArgSerArgProLysSerGlnuLeuAlaLeuSerGlnuArgArgSerArgArgLys 86
Db      387 GCGCGTGGCGCGCGCGCGCGCGCGCTTGTGCGACACCTTCAACCGAGCGCGCGGTG 446
QY      87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
Db      447 AAGCGCAACGAGCGGAGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 506
QY      107 ArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysGlieThrLeuArg 126

```

```

Db      507 CGCAGCGTCCTCCGACCTTCCCGACGACCAAACTCACCMAATGCAAACTCTGGGC 566
Qy      127 PheAlaHisenTyrlleThraAlaLeuThrglnThraLeuArg11AlaAlaSerleu 146
Db      567 TTCGCTTACACTACATCTGGGCGCTCTCCGAGACCTTCGTTGGCGAGCATGCTC 626
Qy      147 -----TyraAlaLeuGluproProAlaProHisCyglYluleuGlYserPro 162
Db      627 CCTCTCCCCCGCTTCCGCGGCGCCCGCGGCC-----CCGAGCCCC 671
Qy      163 GlyYlProProGlYAspTrpGlYserleuTyserPro-----Val 176
Db      672 GGCACGCGACGCGCGTTCGCTGCTCCAGCGGTTCCCGCGCCCTCTGCTGCGCC 731
Qy      177 SerGlnAlaGlySerleuSerProAlaAlaSer----- 187
Db      732 TCCGCTTCGCGGCCGACGACGCGCGCACCTCCGAGACCTGCGCTACGCTCCGAGAC 791
Qy      188 ---LeuGluGluArgProGlYleuLeuGlyAlaThrSerSerAla----- 201
Db      792 GCCCTGCGGGGCTTCCGCGGCTGCGCCCGCGCGCGCGGCGCTCCCTGCGCTAGCCC 851
Qy      202 -----CysLeuSerProGlYserleuAlaPheSer 211
Db      852 TGCCCGTGGTGTCTCCGTCGCCCGCCCGCACCTTCTCC 887

RESULT 7
US-10-004-717-30
; Sequence 30, Application US/10004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-30

Alignment Scores:
Pred. No.: 1,84e-25 Length: 1385
Score: 375.50 Matches: 104
Percent Similarity: 50.00% Conservative: 19
Best Local Similarity: 42.28% Mismatches: 55
Query Match: 33.32% Indels: 69
DB: 14 Gaps: 10

US-09-595-947E-10 (1-214) x US-10-004-717-30 (1-1385)
Qy      26 GluAspGluVal-----ThraYsProThraSerAlaProProSerPro 39
Db      418 GAAGAGAGAGTACTATGCTGCTGGGCTCGGCTTCCCGGCTCCGCGACCTGACCCG 477
Qy      40 -----ThraYsProThraSerAlaProProSerPro 46
Db      478 ATGCTCTCAGCGCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
Qy      47 -----AlaGluAlaGluGlu-----Gly 52

```

```

Db      538 GGGCAGCGTGGAGGCGGAGCCGACGAGGGGTGACGAGGAGCTCGGCGCTCGGTCGCGG 597
Qy      53 GlYcysArgGlyAlaProArgGlyLeuArgAla-----ArgArgGly 66
Db      598 GGTTCGCG-----CCAGGGCGCTGTGGGCGCTGATGACGAGTGCAGCGTGCAGCGCG 651
Qy      67 GlyArgSerArg-----ProlysSerGlnLeuAlaLeuSerIysGlnArg 81
Db      652 TCGGCTACCGGGCGCTCTCCGAGTCCCAAGACGGCGAGAGCGGTGCAGCGATCAG 711
Qy      82 ArgSerArgArgGlyAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
Db      712 AAGACCGCGAGGCTCAAGAGCCCAACACCGCGAGCGCAACCCGATGCACAACTTAACGCC 771
Qy      102 AlaLeuAspAlaLeuArgGlyValLeuProThraPheProAspAspAlaValLeuThrIys 121
Db      772 GCGCTGACGCGCTGCGGAGAGTCTGCTCCACTTCCCGAGAGTCCCAAGCTCAGAG 831
Qy      122 IleGluThraLeuArgPheAlaHisenTyrlleThraAlaLeuThrglnThraLeuArgIle 141
Db      832 ATCAGAGCGTGGCTGCTGCGCCACATTAATCTGGGCGCTCAGCGAATCTGCGCGCTG 891
Qy      142 AlaAspHisSerleuTyraAla-----Leu 149
Db      892 GCGGACCACTGCGCGCGCGCGGTGCGCTCCAGGGGCGCTCTTCAAGAGGCGGTGCTC 951
Qy      150 GluProProAlaProHisCyglYluleuGlySerProGlYlProProGlYAspTrp 169
Db      952 CTGAGCCCGGAGGCTGCGCTGCGCGCGCGCGAGAGAGCCCTTCTCCACTTCTCTCG 1011
Qy      170 GlySerleuTyserProValSerGlnAlaGlySerleuSerPro----- 184
Db      1012 AGCTGCACCAACAGCCCGGCTCATCTCCACTCCAGTCCCATATACGTGCACCTTTA 1071
Qy      185 ---AlaAlaSerleuGluArgProGlYleuLeuGlyAlaThrSerSerAla----- 201
Db      1072 TCGCCCGCTAGCCC-CGGGTCAAGCTGACTGAGAGCCCGACCTCCGAGAGACA 1130
Qy      202 -----CysLeuSerPro 205
Db      1131 TCGTTATCGCGCTCACTT 1148

RESULT 8
US-10-004-717-6
; Sequence 6, Application US/10004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-6

Alignment Scores:
Pred. No.: 5.4e-25 Length: 1412
Score: 370.50 Matches: 106
Percent Similarity: 48.47% Conservative: 21

```

Best Local Similarity: 40.46%  
Query Match: 32.87%  
DB: 14 Gaps: 11

US-09-595-947E-10 (1-214) x US-10-004-717-6 (1-1412)

```

QY 16 GIUthrgluArgSerPheProArgAla-----SerGluAspGluVal 29
DB 73 GAGCGCGGTAGATGTTGCTCAAAATCTGAGACTCTGAGATTGAGAGAGAGAGAGTGA 132
QY 30 -----ThyCysProThrSerAlaProProSerPro----- 39
DB 133 CTGATCTGCTGGCTCGGCTTCCCGGCTCGGAGCCCTGACCCCTGATCTCTCCAGC 192
QY 40 -----ThyArgThrProGlyAsnGly----- 46
DB 193 GCGGACGAGAGAGAGAGAGAGAGAGCTGCCCGGCGGCTCCGCGGCTGGAGCGTGA 252
QY 47 AlaGluAlaGluGlu-----GlyGlyCysArgGly 56
DB 253 GCGGAGCGCGGCGAGGCGGTGACAGGCGAGTCCGCGGTGCGGCGGCTGGCGG--- 309
QY 57 AlaProArgGlyLeuArgAla-----ArgArgGlyGlyArgSerArg 70
DB 310 ---CCAGGCGCGCTGCTGGGCTGTGATGACAGATGCAAGCTGCGGCGGCTGCGGCTCACGG 366
QY 71 -----ProLysSerGluLeuAlaLeuSerGlyAlaArgArgSerArg 85
DB 367 GCGGCTCCCGAGGTGCCAGAGCGGCGGAGACGGGTGACCGCATCAAGAGACCGGAG 426
QY 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
DB 427 CTCAGAGCCCAACACCGGAGCGCAACCGCATGACAACTTAAAGCGCGCTGAGACGG 486
QY 106 LeuArgGlyValLeuProThrPheProAspAspAlaValLeuThrLysIleGluThrLeu 125
DB 487 CTGCGCGAGGTCTGCTGCCACCTTCCCGAGATGCCAAGCTCAAGAGATGACAGCGTG 546
QY 126 ArgPheAlaHisAsnTyrIleThrPheAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
DB 547 CGCTTCGCCCAATTAATCACTGCGGCTCTCACCGAGACTCTGCGCTGCGGAGCACTGCC 606
QY 146 LeuTyrAla-----LeuGluProProAla 153
DB 607 GCGGCGCGGCTGCTGCCAGGCGGCTCTTCAACGAGCGGTGCTCTGAGCCGAG 666
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
DB 667 GCTGCGCTGCGGCGGCGGAGCGGAGCCTTCTTCCCTCTGAGGCTGACCAAC 726
QY 174 SerProValSerGluAlaGlySerLeuSerPro-----AlaAlaSer 187
DB 727 ACCCGGCGCTATCTCTCAACTCTCAAGCTCCCAATACGCTGACCTTATCGCCGCTAAC 786
QY 188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla-----CysLeu 203
DB 787 CC-CGGGTGACAGGTGACTACTGCGAGCCCGACCTCCGAGAGACATCGTTATGCGCC 845
QY 204 SerPro 205
DB 846 TCACCT 851

```

RESULT 9  
US-10-004-717-37

; Sequence 37, Application US/10004717  
; Publication No. US2002019265A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGEN, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P018999US4

; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO: 37  
; LENGTH: 1412  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-717-37

Alignment Scores:  
Pred. No.: 5.4e-25 Length: 1412  
Score: 370.50 Matches: 106  
Percent Similarity: 48.47% Conservative: 21  
Best Local Similarity: 40.46% Mismatches: 61  
Query Match: 32.87% Indels: 75  
DB: 14 Gaps: 11

US-09-595-947E-10 (1-214) x US-10-004-717-37 (1-1412)

```

QY 16 GIUthrgluArgSerPheProArgAla-----SerGluAspGluVal 29
DB 73 GAGCGCGGTAGATGTTGCTCAAAATCTGAGACTCTGAGATTGAGAGAGAGAGAGTGA 132
QY 30 -----ThyCysProThrSerAlaProProSerPro----- 39
DB 133 CTGATCTGCTGGCTCGGCTTCCCGGCTCGGAGCCCTGACCCCTGATCTCTCCAGC 192
QY 40 -----ThyArgThrProGlyAsnGly----- 46
DB 193 GCGGACGAGAGAGAGAGAGAGAGAGCTGCCCGGCGGCTCCGCGGCTGGAGCGTGA 252
QY 47 AlaGluAlaGluGlu-----GlyGlyCysArgGly 56
DB 253 GCGGAGCGCGGCGAGGCGGTGACAGGCGAGTCCGCGGTGCGGCGGCTGGCGG--- 309
QY 57 AlaProArgGlyLeuArgAla-----ArgArgGlyGlyArgSerArg 70
DB 310 ---CCAGGCGCGCTGCTGGGCTGTGATGACAGATGCAAGCTGCGGCGGCTCACGG 366
QY 71 -----ProLysSerGluLeuAlaLeuSerGlyAlaArgArgSerArg 85
DB 367 GCGGCTCCCGAGGTGCCAGAGCGGCGGAGACGGGTGACCGCATCAAGAGACCGGAG 426
QY 86 LysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
DB 427 CTCAGAGCCCAACACCGGAGCGCAACCGCATGACAACTTAAAGCGCGCTGAGACGG 486
QY 106 LeuArgGlyValLeuProThrPheProAspAspAlaValLeuThrLysIleGluThrLeu 125
DB 487 CTGCGCGAGGTCTGCTGCCACCTTCCCGAGATGCCAAGCTCAAGAGATGACAGCGTG 546
QY 126 ArgPheAlaHisAsnTyrIleThrPheAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
DB 547 CGCTTCGCCCAATTAATCACTGCGGCTCTCACCGAGACTCTGCGCTGCGGAGCACTGCC 606
QY 146 LeuTyrAla-----LeuGluProProAla 153
DB 607 GCGGCGCGGCTGCTGCCAGGCGGCTCTTCAACGAGCGGTGCTCTGAGCCGAG 666
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
DB 667 GCTGCGCTGCGGCGGCGGAGCGGAGCCTTCTTCCCTCTGAGGCTGACCAAC 726
QY 174 SerProValSerGluAlaGlySerLeuSerPro-----AlaAlaSer 187
DB 727 ACCCGGCGCTATCTCTCAACTCTCAAGCTCCCAATACGCTGACCTTATCGCCGCTAAC 786

```

QY 188 LeuGIuIaProGIyLeuGIuIaIaThrSerSerAla-----CysLeu 203  
DB 787 CC-CGGGTCAGACGTCGACTACTGGCAGAGCCCGCAGGAGCATCGTTATGCGCC 845  
QY 204 SerPro 205  
DB 846 TCACCT 851  
RESULT 10  
US-08-722-570-12  
Sequence 12, Application US/08722570  
Publication No. US20030044887A1  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Oufu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1527 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-12  
Alignment Scores:  
Pred. No.: 6.53e-25 Length: 1527  
Score: 370.00 Matches: 99  
Percent Similarity: 52.97% Conservative: 17  
Best Local Similarity: 45.21% Mismatches: 63  
Query Match: 32.83% Indels: 40  
DB: 8 Gaps: 8  
US-09-595-947e-10 (1-214) x US-08-722-570-12 (1-1527)  
QY 3 ProGIuIaProGIyLeuGIuIaIaThrSerSerAla-----CysLeu 203  
DB 293 CCAGAGCAACAGCGGAGCGACGTCGTCGACTTCTCAGCAGAGGAGGCTGTGCCA 352  
QY 23 ArgAlaSerGIu-----AspGIuIaIaThrCysProThr-----SerAlaPro 36  
DB 353 GGCTCCAGCGCTTACCTTCCAGGAGGCTGTCCGTCAGCGCCGAGAGCGGCGCCA 412  
QY 37 ProSerPro---ThrArgThr---ProGIyAsnCysAlaGIuIaGIuIaGIyGIyCysA 55  
DB 413 CCTCTCCGGGCGATCGAAGCTTCCCGTGGCCAGGACGAGAGGAGGAGG----- 462

QY 55 rGIyAlaProArgIuLeuArgIuIaArgArgIyIaYArgSerArgProIySerGIuL 75  
DB 463 -----CGGCGGCGAGCGGAGGTCGCGCGGCGGTCGCTCCAGG 502  
QY 75 euAlaLeuSerIyGIuIaArgSerArgIuIaYArgIuIaAsnAspArgIuIaGAsnA 95  
DB 503 CGTCTCTCAGCTCGCTGCGAGGAGCGCTCGCTCAGGAGGAGGAGGAGGAGGAGGAGG 562  
QY 95 rGMeThIaAspLeuAsnSerAlaLeuAspAlaLeuArgGIyValLeuProThrPheProA 115  
DB 563 GTATGATACCTCAACCTGCGCTGCGAGGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGG 622  
QY 115 sPAspAlaLeuThrIySerIleGIuIaThrLeuArgPheAlaHisAsnTyrlaIaAla 135  
DB 623 AGGACACCAAGCTACCAAGATTGAGAGGCTGCGCTTGGCTACAACTACATCTGGGCGC 682  
QY 135 eutHrGIuIaThrLeuArgIleAlaAspHisSerLeu-----TyrA 148  
DB 683 TGGCTGAGACACTGCGCTGCGAGATCAAGGAGCTTCCGCGGAGGAGGAGGAGGAGGAGG 742  
QY 148 lAluGIuIaProArgAlaProHisCysGIyGIuIeGIySerProGIyIaProProGIyA 168  
DB 743 TCCTGCTCCGAGTGTGTCTCCCTGC-----CTGCCGATCCCGAGCCCGGAGGAGG 796  
QY 168 sP-----TyrGIySer-----LeuTyrSerProV 176  
DB 797 ATACAGACTCTCGGGGCTCGGGGCGGCTGCTCCCTCGGCTACTGTGGGCTACGAC 856  
QY 176 alSerGIuIaGIySerLeuSerProAlaAlaSerLeuGIuIaGIuIaGIyGIy 193  
DB 857 TCTCTGACCCGAGTATGCTCGCTCGCTCAGAGACTTCACTTATGCGCGGCT 909

RESULT 11  
US-08-722-570-13  
Sequence 13, Application US/08722570  
Publication No. US20030044887A1  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Oufu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-13



QY 202 sleuserProGlyserLeuAlaPheSerAapPheLeu 214  
DB 941 CTGGCAGCCCTCTGAGCTCAGCTGAGAACCCCTTCAG 977

RESULT 13  
US-08-722-570-15  
Sequence 15, Application US/08722570  
Publication No. US20030044887A1  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RET/RMS  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-15

Alignment Scores:  
Pred. No.: 6,29e-19 Length: 1277  
Score: 304.00 Matches: 86  
Percent Similarity: 57.29% Conservatve: 24  
Best Local Similarity: 44.79% Mismatches: 64  
Query Match: 26.97% Indels: 20  
Gaps: 8

US-09-595-947E-10 (1-214) x US-08-722-570-15 (1-1277)

QY 24 AlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThrPro 43  
DB 238 AGCTGCGAGAGTAG-----CAGCTACACAGCCGACCAAGCCCGCGCTCAGC-- 345

QY 44 GlyAenCyAlaGluAlaGluGlyCysArgGlyAlaProArgLysLeuArgAla 63  
DB 346 -----CACCTGCACAGGAGCGGACCGAGGAGAGAGAACGCCCGCGATGC 393

QY 64 ArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSer 83  
DB 394 AGG-----AGAGCGGAGCGCGCGAGACCGCTGCTG--AAGATCAAGAAAGCC 441

QY 84 ArgArgGlyGlyAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeu 103  
DB 442 CGGCGCGTTAAAGCCAAATACCGGAGAGAAATCGCATGCACACCTGAATATGCGCTC 501

QY 104 AspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGlu 123  
DB 502 GATTCTCTGAGGAGAGTTCTTACCGTCATTATCCGAGAGCCAACTACCAAGATAGAG 561

QY 124 ThrLeuArgPheAlaHisAsnTrpIleTrpAlaLeuThrGlnThrLeuArgIleAlaAsp 143  
DB 562 ACCTTGCCCTTTGCCCAACATACATCTGGGCTCTTACGAAACTTTCGCCGCCGAC 621

QY 144 -----HisLeuLeuThrAlaLeuLubProProAlaProHisCysGlyGlu-----Leu 159  
DB 622 CAGCTGACCGG-ATCTAC--TTCCACCCGACAGACGACCATTTGATACAGACTCTCA 677

QY 160 GlySerProGlyGlyProProGlyAspTrpGlySerLeuLysSerProValSerGlnAla 179  
DB 678 TCCTTCCTCGTA-GCCCTCTCTGAGCTGACGTGCTCCCATCTCCCAACTTTCGACT 736

QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeu--LeuGlyAlaThr 198  
DB 737 CCTCTCCCGACGACCGCTGCGACCTCCGACAGATGATGATGATGATGATGATGATGAT 796

QY 199 SerSerAlaCysLeuSerProGlySerLeuAlaPhe 210  
DB 797 CTGAGCTCGCTTGAACCCCTTCATGCTGCCCTT 832

RESULT 14  
US-10-004-717-43  
Sequence 43, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: YANG, OI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEARBESS,  
FILE REFERENCE: P01899054  
CURRENT FILING DATE: 2002-08-16  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 43  
LENGTH: 1550  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-43

Alignment Scores:  
Pred. No.: 8,11e-13 Length: 1550  
Score: 238.50 Matches: 78  
Percent Similarity: 39.29% Conservatve: 21  
Best Local Similarity: 30.95% Mismatches: 72  
Query Match: 21.16% Indels: 81  
Gaps: 8

US-09-595-947E-10 (1-214) x US-10-004-717-43 (1-1550)

QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36  
DB 106 AGAAAGTTTGCTAGCAATGTGAGACCAAGAAATTAAGAAACAGAGGCTTTCCA 165

QY 37 -----ProSerProThrArgTrpProGlyAsnCyAlaGluAla 49  
DB 166 AAACAAGTTTCCTTCAGAGAAAGACATTAAGAGGCCCTCGAGAGAAAGCAAGAAA 225

QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59  
DB 226 GAAAGAGAGAAAGAAAGACAGAGAGAGAGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 267

```

QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db 268 ---TTGTCAGAGAGGAGGGGCTCAGAGAAAAAGACCAACCAACTGACATGGGAAAG 324
QY 80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db 325 GTC---AAGTTCAGAGAGAGAGAGCTAATGCGCGAGAGAGACCGAGTGCACGGCCCTC 381
QY 100 AenSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db 382 AATGATGCTCTGGACAACTTTGGGAAAGGTGCTCCCTGTTACTCTTAAACCAAAAGCTG 441
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrTleTyrAlaLeuThrGlnThrLeu 139
Db 442 TCCAAATATGAAACTTTACGACTGGCCAAAATTAATCATCTGGGCACTTTCTGAAATCTG 501
QY 140 ArgIle----- 141
Db 502 AGGATTTGGCAGAGACCGGATCTGCTCAGCTTGTCCAAAACCTTATGCAAGGCTTTCC 561
QY 141 ----- 141
Db 562 CAGCCAACTACAAACTTGTGGAGGCTGCTTACAGCTCAGCCAGAGTTTCTGATG 621
QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProProAla 153
Db 622 GGTCAAGGTTGGGAGAGCTGCCACACACAGAGTCACTTACTCTTACCATTTACCAACCC 681
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
Db 682 TACCAAGCCCTGAGCTGGCCACCTCCCAAGG-----CAT 717
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
Db 718 GGGACTCTTGATTAATTCAGAGTCCATGAAGAACCTTACATTAATGATGATGATTC 777
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db 778 TTCTATGAAGTACCTCCCTGAGTGTGCCAGCCCT 813

```

## RESULT 15

```

US-10-004-717-8
; Sequence 8: Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHEI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1957
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-8

```

## Alignment Scores:

```

Pred. No.: 1,04e-12 Length: 1957
Score: 238.50 Matches: 78
Percent Similarity: 39.29% Conservative: 22
Best Local Similarity: 30.95% Mismatches: 72
Query Match: 21.16% Indels: 81

```

```

DB: 14 Gaps: 8
US-09-595-947e-10 (1-214) x US-10-004-717-8 (1-1957)
QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36
Db 192 AGAAAGTTTCTAGCAATGTGAGAGACCAAGAAACAAATTAAGAAACGAGAGCTTTTCA 241
QY 37 -----ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
Db 242 AAACAAAGTTTCTTTCGAGAAAGACATTAAGAGGCTTGGAGAGAAACGAGAA 301
QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59
Db 302 GAAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGATGGC----- 343
QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db 344 ---TTGTCAGAGAGAGAGGGGCTCAGGAAAAAGACCAACCAACTGACATGGGAAAG 400
QY 80 GlnArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db 401 GTC---AAGTTCAGAGAGAGAGAGAGCTAATGCGCGAGAGAGACCGGATGCAACGGCCCTC 457
QY 100 AenSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db 100 AATGATGCTCTGGACAACTTTGGGAAAGGTGCTCCCTGTTACTCTTAAACCAAAAGCTG 517
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrTleTyrAlaLeuThrGlnThrLeu 139
Db 518 TCCAAATATGAAACTTTACGACTGGCCAAAATTAATCATCTGGGCACTTTCTGAAATCTG 577
QY 140 ArgIle----- 141
Db 578 AGGATTTGGCAGAGACCGGATCTGCTCAGCTTGTCCAAAACCTTATGCAAGGCTTTCC 637
QY 141 ----- 141
Db 638 CAGCCAACTACAAACTTGTGGAGGCTGCTTACAGCTCAGCCAGAGATTTCCTGATG 697
QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProProAla 153
Db 698 GGTCAAGGTTGGGAGAGCTGCCACACACAGAGTCACTTACTCTTACCATTTACCAACCC 757
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
Db 758 TACCAAGCCCTGAGCTGGCCACCTCCCAAGG-----CAT 793
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
Db 794 GGGACTCTTGATTAATTCAGAGTCCATGAAGAACCTTACATTAATGATGATGATTC 853
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db 854 TTCTATGAAGTACCTCCCTGAGTGTGCCAGCCCT 889

```

Search completed: February 2, 2004, 21:41:42  
 Job time : 361 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 2, 2004, 17:09:49 / Search time 2044 Seconds  
(without alignments)  
2544.598 Million cell updates/sec

Title: US-09-595-947E-10  
Perfect score: 1127  
Sequence: 1 MTPQSGAPVQVTRTERS.....LGATSSACISPGSLAFSDPL 214

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODU=frame+.p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO\_SPOOL/US095594/rumat\_02022004\_154934\_8059/app\_query.fasta\_1.391  
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human0.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=tbl -NORM=ext -HEADPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0955947@cgn2.1 1.2810 @rumat\_02022004\_154934\_8059 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-GAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : EST.\*  
1: em\_estda.\*  
2: em\_estchum.\*  
3: em\_estlu.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hic.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hic.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estum.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vit.\*  
21: em\_gss\_fur.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rnd.\*  
26: em\_gss\_ping.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*  
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	809	71.8	1540	11 AK008017	AK008017 Mus muscu
2	726	64.4	1025	16 BY708009	BY708009 Mus muscu
3	676.5	60.0	593	28 A2296526	A2296526 RPICT-23-1
4	340.5	30.2	932	14 CA979119	CA979119 AGENCOURT
5	327	29.0	865	14 CA96091	CA96091 AGENCOURT
6	327	29.0	1098	13 BQ169355	BQ169355 STR00303
7	318	28.2	603	14 CD282259	CD282259 G39271.16
8	317.5	28.2	632	12 BQ101277	BQ101277 BQ101277
9	314	27.9	687	14 CA945402	CA945402 UI-M-FPO-
10	311	27.6	604	13 BX308104	BX308104 BX308104
11	310.5	27.6	600	12 BG808248	BG808248 2083-52 M
12	299	26.5	600	13 BG924937	BG924937 7103-91 M
13	298.5	26.5	1001	9 AL540071	AL540071 AL540071
14	296.5	26.3	1039	10 BE780690	BE780690 601469349
15	295	26.2	730	13 BU612495	BU612495 UI-M-FPO-
16	285.5	25.3	1037	29 CNS03VJ9	AL262494 Tetradon
17	281	24.9	823	13 BU054481	BU054481 UI-M-FPO-
18	280	24.8	629	12 BU093114	BU093114 BU093114
19	278	24.7	595	12 BU030202	BU030202 BU030202
20	267.5	23.7	401	9 AW147434	AW147434 da02h12.Y
21	265	23.5	947	13 BX419330	BX419330 BX419330
22	251	22.3	588	9 AV673464	AV673464 AV673464
23	249	22.1	490	9 AV995230	AV995230 AV995230
24	249	22.1	711	13 BW275045	BW275045 BW275045
25	243	21.6	814	13 BQ178789	BQ178789 UI-M-EVO-
26	239	21.2	1003	29 CANS021B1	AL198694 Tetradon
27	234.5	20.8	1038	13 BX419494	BX419494 BX419494
28	231	20.5	1022	13 BU113216	BU113216 603129939
29	229.5	20.4	932	13 BU138911	BU138911 603132642
30	228	20.2	704	13 BU057851	BU057851 UI-M-FPO-
31	228	20.2	740	14 CA319439	CA319439 UI-M-FPO-
32	228	20.2	835	14 CA320553	CA320553 UI-M-FPO-
33	226.5	20.1	710	13 BU057238	BU057238 UI-M-FPO-
34	226.5	20.1	742	13 BQ572268	BQ572268 UI-M-FPO-
35	226.5	20.1	770	13 BQ572426	BQ572426 UI-M-FPO-
36	224.5	19.9	688	10 BG699059	BG699059 602678696
37	224.5	19.9	711	2 HSM067528	Bx478059 Homo sapi
38	224.5	19.9	901	13 BQ424098	BQ424098 AGENCOURT
39	224.5	19.9	1967	11 BC022560	BC022560 Homo sapi
40	222.5	19.7	953	9 AU067624	AU067624 AU067624
41	221.5	19.7	781	13 BU611678	BU611678 UI-M-FPO-
42	221.5	19.7	872	13 BX453565	BX453565 BX453565
43	219.5	19.5	724	10 BE783567	BE783567 601471617
44	219	19.4	722	2 HSM078917	Bx508253 Homo sapi
45	219	19.4	875	29 CANS02BKP	AL189970 Tetradon

## ALIGNMENTS

RESULT 1  
AK008017  
LOCUS  
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length  
(Drosophila), full insert sequence.  
ACCESSION AK008017  
VERSION AK008017.1 GI:12841941  
KEYWORDS HTG; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
10349636

1. Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
1042159

2. Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861

3. Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, S., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11076861

4. Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Aizawa, T., Hara, A., Fukunishi, Y., Komoto, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L. M., Staudli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balardini, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombauts, P., Nordone, P., Rind, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H., Toyooka, K., Wang, K. H., Weitz, C., Whitaker, C., Wilming, L., Wrynha-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S. and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11217851

5. The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
11217851

6. (bases 1 to 1540)  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furumoto, K., Furuno, M., Hasegawa, T., Hara, A., Hayatsu, N., Hiramoto, Y., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Teijima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.  
Direct Substitution  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome

## COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: <http://genome-gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer (5' GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTN 3'), cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTN 3'. cDNA was cleaved with XhoI and SclI. Cloning sites, 5' end: XhoI; 3' end: SclI.

Host: SOLR  
Location/Qualifiers

## FEATURES

## source

1..1540  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:2010001M19"  
/db\_xref="MGI:1907403"  
/db\_xref="taxon:10090"  
/clone="2010001M19"  
/sex="male"  
/tissue\_type="small intestine"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"

## CDS

241..885  
/note="unnamed protein product; atonal homolog 5 (Drosophila) [MGD:MGI:893591, GB|NM\_009719, evidence: BLASTN, 99%, match=648]  
putative"  
/codon\_start=1  
/protein\_id="BAB25411.1"  
/db\_xref="GI:12841942"  
/db\_xref="MGI:893591"  
/translation="MAAPPLDALATIQVPEPTQPPGASDHEVLSNSPTPSHTLIPDCSEAEVDDCGTSLKLRAGGRPRPSESLKRSRRKARDRENRHNSALDALAGVLPPTPDADKLTKEITRPAHVIALTQTLAIHSLYGPPEPVCGCLSGPEGSGNMGWSTYSPVSGANISPTASLEFPGLQVPSSTLPLGALVFSDFL"

BASE COUNT 306 a 485 c 397 g 352 t  
ORIGIN

## Alignment Scores:

Pred. No.:	8,47e-46	Length:	1540
Score:	809.00	Matches:	161
Percent Similarity:	82.33%	Conservative:	16
Best Local Similarity:	74.88%	Mismatches:	36
Query Match:	71.78%	Indels:	2
DB:	11	Gaps:	2

US-09-595-947E-10 (1-214) x AK008017 (1-1540)

QY	1	MeatnProGlnInProSerGlyAlaProThrValGlnValThrArgGluThrGlnArgSer	20
DB	241	ARGGGCCCTCACCCTTGAGATGCGCTACACCAACAGATGTCCAGAGACACACACT	300
QY	21	PheProArgAlaSerGlnuSpGlnValThrCysProThrSerAlaProProSerProThr	40
DB	301	TTTCCCGAGAGCTCGACACACAGAGTGCCTCAATTCACACCCACTGACACACT	360
QY	41	ArgThrProGlyAsnGlyAlaGluGluGluGlyGlyCysArgGlyAlaProArgGly	60
DB	361	CTCATACCTAGGAGCTCTCGAACACAGAGTGGTACTGCGAGGAGACTTCGAGGAAG	420
QY	61	LeuArgAlaArgArgGlyValArgSerArgProGlySerGluLeuAlaLeuSerArg	80

FEATURES	SOURCE	LOCATION/Qualifiers
BASE COUNT	215 a	316 c 286 g 207 t 1 others
ORIGIN		
Alignment Scores:		
Pred. No.:	2,75e-40	Length: 1025
Score:	726.00	Matches: 160
Percent Similarity:	82.33%	Conservative: 17
Best local Similarity:	74.42%	Mismatches: 36
Query Match:	64.42%	Indels: 7

```

DB:      14      Gaps:      2
US-09-595-947E-10 (1-214) x BY708009 (1-1025)
QY      1 MettPrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB      241 ATGGGGCCCTCATCCCTTGGATGGCTGACACATCCCAAGTGTCCCAAGACACAAACCT 300
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
DB      301 TTTCGGAGCCTCGGACCAAGTGTCTCAATTCATCCACCCACTTGCACACT 360
QY      41 ArgThrProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgGlyS 60
DB      361 CTCATACCTAGGAGACTGCTCCGAGCAGAAAGTGGTGACTGCGCAGAGGACTCGAGGAG 420
QY      61 LeuArgAlaArgArgGlyGlyValArgSerArgProGlySerGluLeuAlaLeuSerGlyS 80
DB      421 CTCGGGCGCCGACGCGGAGGCGGCAACAGGCCCAAGAGAGTGGCACTCGACCAAAAG 480
QY      81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB      481 CCAAGAGCCCGGCGCAAGAGGCCAATGATCGGAGCGCAATGCAATGCAACCTCAAC 540
QY      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB      541 TCGGGCGTGGATGCGGTGCGGCGGTCTCTGCCCACTTCCGGATGACGCCAAACTTACA 600
QY      121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGluThrLeuArg 140
DB      601 AAGTCGAGACCCCTGCTGCGCCCACTACATCTGAGGCACTGACTCAGACGCTGCGC 660
QY      141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyGluLeuGly 160
DB      661 ATAGCGGACCAACGCTGTATGAGCCCGGAGCCCCCTGTGCC--TGTGAGAGCTGGGG 717
QY      161 SerProGly---GlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
DB      718 AC-CGCGGACGTGGCTCCCAAGCGGACTGGGGCTGATATCATCTCCATCTCCCAAGG 776
QY      180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB      777 GGTAACCTAGGCCACAG-CGCTCATTTGAGAGATTA-CCTGGCTTGA-CGATGCCACTCC 833
QY      200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB      834 CCACTCAT-CTGCTCCCGGAGCACTGTGTCTCAGACTTCTG 877

RESULT 3
LOCUS   AZ296526      593 bp      DNA      linear      GSS 27-JUL-2000
DEFINITION RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18
ACCESSION AZ296526
VERSION   AZ296526.1 GI:9538311
KEYWORDS GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
Zhao, S., Nieman, W., Feidlyum, T., Malek, J., Shatsman, S., Akinrit,
B., Levins, M., Megam, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
Other GSSs: RPCI-23-160G18.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

```

```

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html
Plate: 160 row: G column: 18
Seq primer: 17
Class: BAC ends.

FEATURES
    source          location/Qualifiers
    1..593
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPCI-23-160G18"
    /sex="Female"
    /lab_host="DH10B"
    /clone_1ib="RPCI-23"
    /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
    EcorI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or
    brain genomic DNA was isolated and partially digested
    with a combination of EcorI and EcorI Methase. Size
    selected DNA was cloned into the pBACe3.6 vector at the
    EcorI sites. The ligation products were transformed into
    DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      88 a      159 c      213 g      133 t
ORIGIN
Alignment Scores:
Pred. No.:      3,96e-37      Length:      593
Score:          676.50      Matches:      136
Percent Similarity: 82.68%      Conservative: 12
Best Local Similarity: 75.98%      Mismatches: 30
Query Match:    60.03%      Indels:      2
DB:             28      Gaps:      1
US-09-595-947E-10 (1-214) x AZ296526 (1-593)
QY      3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
DB      547 CCTCATCCCTTGGATGGCTGCTCACTCCATCCAGTGTCCCAAGACACAAACCTTTTCCC 488
QY      23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThr 42
DB      487 GGAGCTCGAGCACCAAGTGTCTCAATTCATCCACCCCACTTACCCCACTTCATTA 428
QY      43 ProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgGlySerLeuArg 62
DB      427 CCTAGGAGACTGCTCCGAGGACCAAGTGGTGACTGCCAGGGAGCTCGAGAGAGTCCGC 368
QY      63 AlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerGlyGlnArg 82
DB      367 GCCCGACGCGGAGCGGACACAGCGCCCAAGAGCGAGTTGGCACTCAGCAAAACAGGAGA 308
QY      83 SerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102
DB      307 ACCCGCGCAAAAGCCCAATGATGGAGCGCAATCGCATCGCAACCTCAACTCGGGG 248
QY      103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122
DB      247 CTGATGCGCTGCGGGGTCTCTGCTCCACTTCCGGATGACGCCAACTTCAAAAGATC 188
QY      123 GluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGluThrLeuArgIleAla 142
DB      187 GAGACCTCGGCTTGGCCACACATACATCTGGCAGTCACTCAGACGCTGGCAGATAGG 128
QY      143 AspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyGluLeuGlySerPro 162
DB      127 GACCAAGCTTTATAGGCCCGGAGGCCCTGTGCTCTG---GAGAGCTGGGGAGCCCC 72
QY      163 GlyGlyProProGlyAspTyrPlySerLeuTyrSerProValSerGlnAlaGlySer 181

```

Db 71 GGAGGGGCTCCACGCGGAGCTGGGCTCATCACTCCCAAGTCTCCCAAGCGGGAACC 15  
 RESULT 4 CA979119 932 bp mRNA linear EST 06-JAN-2003  
 LOCUS CA979119  
 DEFINITION AGENCOURT 11295215 NIH MGC 164 Mus musculus cDNA clone  
 IMAGE:30146192 5', mRNA sequence.  
 CA979119  
 ACCESSION CA979119.1 GI:27511773  
 VERSION  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 C ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgi.nci.nih.gov/  
 1 (bases 1 to 932)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Dr. David Rowe and Dr. Mina  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
 http://image.llnl.gov  
 Plate: NDAW0061 row: h column: 09  
 High quality sequence start: 16  
 High quality sequence stop: 640.  
 Location/Qualifiers  
 1..932  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30146192"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_11b="NIH MGC 164"  
 /note="Vector: pCMV-SPORT6.1.cdb, Site\_1: EcoRV, Site\_2:  
 NotI; Non-normalized full-length enriched library from  
 pooled mouse embryonic limb, maxilla and mandible, day  
 10.5 and 11.5 (size selected for the 0.5-1 kb fragments)  
 Cloned directionally, priming method: Oligo-dT. cDNA  
 enrichment: >1k bp. Average insert size 1.8k bp. Priming  
 sequence: 5'-GACTAGTCTAGATCGGAGCGGCGCC(T) 3'. Tissue  
 contributed by, David Rowe. Library constructed by Resgen,  
 Invitrogen Corp."  
 BASE COUNT 153 a 349 c 276 g 153 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 6.1e-14 Length: 932  
 Score: 340.50 Matches: 103  
 Percent Similarity: 50.00% Conservative: 21  
 Best Local Similarity: 41.53% Mismatches: 64  
 Query Match: 30.21% Indels: 61  
 DB: 14 Gaps: 11  
 US-09-595-947E-10 (1-214) x CA979119 (1-932)  
 QY 4 GlnProSerGlyAlaProThrValGlnValThr---ArgGluThrGlnArgSerPhePro 22  
 Db 208 AGGCGGAGCTCCAAACCTCTGCTCGCTGCGCTCTGCACACCTGCAGAGATGCTGCCC 267  
 QY 23 -----ArgAlaSerGlnAspGluValThrCys 31  
 Db 268 CTTTGAGACGCTGCATCTGCATCTGCATCTGCCAGCAGCAAGCAGCAGCAGCTT 327  
 QY 32 ProThrSerAlaProProSerProThrArgThrProGlyAsnGlyAlaGluGluGlu 51  
 Db 328 CCAGCTTCTCAACGAGAGAGACT---GTGCCAGGCTACAGCCCTTA-GCTTCACCC 383

QY 52 GlyIlyCys-----ArgGlyAlaPro----- 58  
 Db 384 TCGGGGCTGTCCGTCCAGCCCGGAGAGCGCTCCCGCCCTCTCCGGGGCATTCAGTGT 443  
 QY 59 -----ArgIlySerGlnAlaArgArgGlyIlyArgSerArgProIlys 72  
 Db 444 CCCGGTCCAGAGCAG 503  
 QY 73 SerGluLeuAlaLeuSerIlySerGlnArgArgSerAlaArgIlyIlyAlaAsnAspArgGlu 92  
 Db 504 TCCGAGGCTCTGCTCCTCCTCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 563  
 QY 93 ArgAnaArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThr 112  
 Db 564 CGCAACCGGATGACACACTTCAACCTGCGCTGAGAGCGCTTCCGAGAGCTGTGCTCC 623  
 QY 113 PheProAspAspAlaIlyLeuThrIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIlyIly 132  
 Db 624 TTTCCCGAGAGACACCAAGCTGACCAAGATTGAGCGTGGCGCTTCCCTACAACTACAT 683  
 QY 132 eTTP-AlaLeuThrGlnThrLeuAlaGlyIleAlaAspHisSerLeuTyralaLeuGluProp 152  
 Db 684 CTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 728  
 QY 152 roAlaProHisCysGlyIlyLeuGlySerProGlyIlyProProGlyAspTyrGlySerL 172  
 Db 729 CCGGGGGC-----AGTCCCGGGAGAGCGCTCT-----GCTCCG 764  
 QY 172 euTyrSerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgP 192  
 Db 765 CAGTGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803  
 QY 192 roGlyLeuLeu-----GlyAlaThrSerSerAlaCysLeuSerPro-----G 206  
 Db 804 CAGCCACACTGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 863  
 QY 206 IySerLeuAlaPheSerAsp 212  
 Db 864 GGCAATCACCACTTCTGTGAC 883  
 RESULT 5 CA496091 865 bp mRNA linear EST 14-NOV-2002  
 LOCUS CA496091  
 DEFINITION AGENCOURT 10812072 NCI\_GGA\_Zemb2 Dario rerio cDNA clone  
 IMAGE:5789795 5', mRNA sequence.  
 CA496091  
 ACCESSION CA496091.1 GI:24959169  
 VERSION  
 KEYWORDS EST.  
 SOURCE Dario rerio (zebrafish)  
 ORGANISM Dario rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 NIH-MGC http://mgi.nci.nih.gov/  
 1 (bases 1 to 865)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Leonard I. Zon, M.D.  
 cDNA Library Preparation: Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
 http://image.llnl.gov  
 Plate: LHAM14290 row: n column: 02  
 High quality sequence stop: 658.  
 Location/Qualifiers  
 1..865  
 /organism="Dario rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"

```

/clone="IMAGE:6789795"
/cisue_type="embryo"
/lab_host="DH10B (71-resistant)"
/clone_lib="NCI CGAP Zemb2"
/notes="Vector: pCMV-SPORT6.ccdB, Site 1: EcoRV, Site 2:
NotI; Cloned unidirectionally. Primer: Oligo dt. Average
insert size 2 kb. Constructed by J. Wang (Research
Genetics, Invitrogen Corp) from tissue donated by L. Zon
(Harvard University). Note: this is a NCI CGAP library."

```

BASE COUNT 226 a 256 c 214 g 169 t

ORIGIN

#### Alignment Scores:

	4.77e-13	Length:	865
Pred. No.:	327.00	Matches:	84
Score:	53.14%	Conservative:	26
Percent Similarity:	40.58%	Mismatches:	65
Best Local Similarity:	29.02%	Indels:	32
Query Match:	14	Gaps:	6

US-09-595-947E-10 (1-214) x CA496091 (1-865)

```

QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13
DB 85 CCTCAAGCATCTCCAGCCCAATAGATTATCAACATGAGATCTGTAATCTCCGAT 144
QY 14 -----ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
DB 145 ATGGAACCTCAAGCTGACCTACTCTTTTCCGAC---ACGATGATGAAGACTCGGCC 201
QY 32 ProThrSerAlaProSerProThrArgThrProGlyAsnCysAlaGluValGluGlu 51
DB 202 ACAGAGCTCCAGCCCGCTCCCGCGG-----TCTCTCTGC----- 237
QY 52 GlyGlyCysArgGlyAlaProArgGlyLeuArgAlaArgArgGlyValArgSerArgPro 71
DB 238 GGAAAAACCACTGCGCTCCAGCCCGGCTCCAGCAAGAAAAAGCGCGGCGCGCG 297
QY 72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysAlaAsnAspArg 91
DB 298 AGGAACGAAACCACTGTCACGTCGTAAGAAAGAACCGAGCTGAGGCCAAGCAGCC 357
QY 92 GluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
DB 358 GAGAGGAAACAGATGACCAACCTTAACGACGATTGATGCTTGAAGAGCGCTCGCT 417
QY 112 ThrPheProAspAspAlaLeuLeuThrLysIleGluThrLeuArgPheAlaHisAsnThr 131
DB 418 GCGTTTCTGACGACCAAGCTGACCAAAATTGAGACTCTGCGCTTCCCTCAACACTAC 477
QY 132 IleTPAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu----- 146
DB 478 ATCTGGGACCTTTGGAGACCATCCGGATCGCAGACCAAGAGCGCAAGTCAGAGAC 537
QY 147 ---TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyPro 165
DB 538 GGTGCGCTCTGCTCCCGAGACTAAGCTGATGCGAGATGACCCAGCCCGGCAAGTAC 597
QY 166 ProGlyAspTyrGlySerLeuThrLysSerProValSerGlnAlaGlySerLeu----- 182
DB 598 TCTTGTCTCTGGCGGCGGCGGCGATCTGCTCTTCAACCGCTTACTGCAACTCA 657
QY 183 -----SerProAlaAla 186
DB 658 GACCCGGGACCCCGCGAGCC 678

```

RESULT 6  
 B0169355 1098 bp mRNA linear EST 01-MAY-2002  
 LOCUS STR00303 segmentation stage cDNA library Dario reio cDNA clone  
 DEFINITION CB260 5' similar to NEUROGENIN 1, mRNA sequence.  
 ACCESSION B0169355  
 VERSION B0169355.1 GI:20376783

KEYWORDS  
 SOURCE  
 ORGANISM

EST.  
 Dario reio (zebrafish)  
 Dario reio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes  
 Cyprinidae; Dario.  
 1 (bases 1 to 1098)  
 Loppin,B., Pflumlo,S., Steffan,T., Heyer,V., Furchauer,M., Thisse  
 ,C. and Thisse,B.  
 Expression of the zebrafish genome during embryogenesis (2002)

TITLE  
 JOURNAL  
 COMMENT

Contact: Thisse B  
 Institut de Genetique et de Biologie Molculaire et Cellulaire  
 CNRS, INSERM, ULP  
 1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex  
 , France  
 Tel: 33 3 88 65 32 60  
 Fax: 33 3 88 65 32 01  
 Email: thisse@igmc.u-strasbg.fr  
 EST from a cDNA of a gene whose expression is spatially restricted  
 during embryogenesis. We have established its expression pattern  
 during embryonic development by whole mount in situ hybridization  
 on zebrafish embryos from the gastrula stage to 2 days of  
 development. The corresponding data are available on the zebrafish  
 community database at <http://zfinfo.org/cDNA> library preparation: B.  
 Riggelman. DNA Sequencing by: IGBMC sequencing facility. Clone  
 distribution: zebrafish international resource center at the  
 University of Oregon (Institute of Neuroscience, 1254 University of  
 Oregon, Eugene, OR 97403-1254)  
 Seq primer: 73 ATTAACCTCACTAAAGGA.  
 Location/Qualifiers

FEATURES  
 source  
 1..1098  
 /organism="Dario reio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="CB260"  
 /dev\_stage="segmentation stage embryos"  
 /clone\_lib="segmentation stage cDNA library"  
 /note="Vector: Lambda zap; Site 1: EcoRI; Site 2: XhoI;  
 Oligo dt cDNA library constructed from RNA pooled from  
 segmentation stage zebrafish embryos"

#### BASE COUNT

284 a 296 c 264 g 254 t

ORIGIN

#### Alignment Scores:

	6e-13	Length:	1098
Pred. No.:	327.00	Matches:	84
Score:	53.14%	Conservative:	26
Percent Similarity:	40.58%	Mismatches:	65
Best Local Similarity:	29.02%	Indels:	32
Query Match:	13	Gaps:	6

US-09-595-947E-10 (1-214) x B0169355 (1-1098)

```

QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13
DB 115 CCTCAAGCATCTCCAGCCCAATAGATTATCAACATGAGATCTGTAATCTCCGAT 174
QY 14 -----ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
DB 175 ATGGAACCTCAAGCTGACCTACTCTTTTCCGAC---ACGATGATGAAGACTCGGCC 231
QY 32 ProThrSerAlaProSerProThrArgThrProGlyAsnCysAlaGluValGluGlu 51
DB 232 ACAGAGCTCCAGCCCGCTCCCGCGG-----TCTCTCTGC----- 267
QY 52 GlyGlyCysArgGlyAlaProArgGlyLeuArgAlaArgArgGlyValArgSerArgPro 71
DB 266 GGAAAAACCACTGCGCTCCAGCCCGGCTCCAGCAAGAAAAAGCGCGAGCGCGCGCG 327
QY 72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArgLysAlaAsnAspArg 91
DB 328 AGGAACGAAACCACTGTCACGTCGTAAGAAAGAACCGAGCTGAGGCCAAGCAGCCG 387

```

QY 92 GluArgAsnArgMetHisAspLeuLeuSerAlaLeuAspAlaLeuArgGlyValLeuPro 111  
 |||||  
 DB 368 GAGAGAACGATGACCACTTAAACGACGATTCGATCTTGAAGAGCTCTGCT 447  
 |||||  
 QY 112 ThrPheProAspAspAlaLeuLeuThrIleGlyValLeuArgPheAlaHisAsnTyr 131  
 |||||  
 DB 448 GCGTTCTCGACGACCAAAAGCTGACCAAAATTGAGACTCTGCTGCTCACAACCTAC 507  
 |||||  
 QY 132 IleTAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu----- 146  
 |||||  
 DB 508 ATCTGGGACCTTTCGAGACCATCCGATCGCAGACCAAGACAGGCGAAGTCAAGAGAC 567  
 |||||  
 QY 147 --TyrAlaLeuGluProProAlaProHisCysGlyLeuLeuGlySerProGlyLeuPro 165  
 |||||  
 DB 568 GGTCCGCTGCTCTCCCGGACTAGCTGATGAGAGACCAAGAGAGAGAGAGAGAGAG 627  
 |||||  
 QY 166 ProGlyAspTyrGlySerLeuTyrSerProValSerGlnAlaGlySerLeu----- 182  
 |||||  
 DB 628 TCTTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 687  
 |||||  
 QY 183 -----SerProAlaAla 186  
 |||||  
 DB 688 GACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 708  
 |||||  
 RESULT 7  
 LOCUS CD282259 603 bp mRNA linear EST 23-MAY-2003  
 DEFINITION G39271.16 NCI CGAP\_Zemb2 Danio, rerio cDNA clone IMAGE:6521499 5',  
 mRNA sequence.  
 ACCESSION CD282259  
 VERSION CD282259.1 GI:31060035  
 KEYWORDS EST.  
 SOURCE Danio rerio (zebrafish)  
 ORGANISM Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes  
 ; Cyprinidae; Danio.  
 1 (bases 1 to 603)  
 Amundsen, C., Cachuela, N., Chen, F., Cheung, L. M., Chong, A., Murray, L.,  
 Oliva, J., Park, C., Reyes, J., Yung, J., and Swimmer, C.  
 Expressed sequence tags from NCI CGAP\_Zemb2, a Danio rerio  
 embryonic library  
 Unpublished  
 Contact: Chen F.  
 JOURNAL  
 COMMENT Exelixis, Inc.  
 170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA  
 Tel: 650 837 7000  
 Fax: 650 837 8300  
 Email: fchen@exelixis.com  
 DNA Sequencing by: Exelixis, Inc. Clone distribution information  
 can be found through the I.M.A.G.E. Consortium/BLN at:  
 http://image.llnl.gov  
 Plate: 14108 row: O column: 3  
 High quality sequence scop: 603.  
 FEATURES  
 location/Qualifiers  
 1..603  
 /organism="Danio rerio"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:7955"  
 /clone="IMAGE:6521499"  
 /tissue\_type="embryo"  
 /lab\_host="DH10B (T1-resistant)"  
 /clone\_lib="NCI CGAP\_Zemb2"  
 /note="Vector: PCMV-SPORT6.cdb; Site\_1: EcoRV; Site\_2:  
 NotI; Cloned unidirectionally. Primer: Oligo dT. Average  
 insert size 2 kb. Constructed by J. Wang (Research  
 Genetics, Invitrogen Corp) from tissue donated by L. Zon  
 (Harvard University). Note: this is a NCI CGAP library."  
 BASE COUNT 159 a 185 c 155 g 104 t  
 ORIGIN  
 Alignment Scores:

Pred. No.: 1.39e-12 Length: 603  
 Score: 318.00 Matches: 76  
 Percent Similarity: 54.89% Conservative: 25  
 Best Local Similarity: 41.30% Mismatches: 57  
 Query Match: 28.22% Indels: 26  
 DB: 14 Gaps: 5  
 US-09-595-947e-10 (1-214) x CD282259 (1-603)  
 QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13  
 |||||  
 DB 79 CCTCAAGATCTCTCCAGCCCAACCAATAGATTATCAATGAGATGATCTACTCCGAT 138  
 |||||  
 QY 14 -----ThrArgGluThrGlnArgSerPheProArgAlaSerGluAspGluValThrCys 31  
 |||||  
 DB 139 ATGAAACCTCAAGCTGATCTACTCTTTCGAC--AGGAGATGAGAGACTCGGC 195  
 |||||  
 QY 32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCyAlaGluAluGlu 51  
 |||||  
 DB 196 AGCAGCTCCACCCCGGCGTCCCGCG-----TCTCTCTGC----- 231  
 |||||  
 QY 52 GlyGlyCysArgGlyAlaProArgLeuAlaArgAlaArgGlyValArgSerArgPro 71  
 |||||  
 DB 232 GGAAGAACCACTGCTCTCCAGCGGCTCCAGCAGAGAAAGAGCGAGGCGCGCG 291  
 |||||  
 QY 72 LysSerGluLeuAlaLeuSerLeuGlnArgArgSerArgGlyValAlaAspArg 91  
 |||||  
 DB 232 AGAAGCAACCACTGCTCTCCAGCGGCTCCAGCAGAGAAAGAGCGCTGAGGCGAGCGC 351  
 |||||  
 QY 92 GluArgAsnArgMetHisAspLeuLeuSerAlaLeuAspAlaLeuArgGlyValLeuPro 111  
 |||||  
 DB 352 GAGAGAACGATGACCACTTAAACGACGATTCGATCTTGAAGAGCTCTGCT 411  
 |||||  
 QY 112 ThrPheProAspAspAlaLeuLeuThrIleGlyValLeuArgPheAlaHisAsnTyr 131  
 |||||  
 DB 412 GCGTTCTCGACGACCAAAAGCTGACCAAAATTGAGACTCTGCTGCTCACAACCTAC 471  
 |||||  
 QY 132 IleTAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu----- 146  
 |||||  
 DB 472 ATCTGGGACCTTTCGAGACCATCCGATCGCAGACCAAGACAGGCGAAGTCAAGAGAC 531  
 |||||  
 QY 147 --TyrAlaLeuGluProProAlaProHisCysGlyLeuLeuGlySerProGlyLeuPro 165  
 |||||  
 DB 532 GGTCCGCTGCTCTCCCGGACTAGCTGATGAGAGACCAAGAGAGAGAGAGAGAGAG 591  
 |||||  
 QY 166 ProGlyAspTyr 169  
 |||||  
 DB 592 TCTTCTCTCTG 603  
 |||||  
 RESULT 8  
 LOCUS BU010277 632 bp mRNA linear EST 05-DEC-2001  
 DEFINITION BU010277 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA142H01 5',  
 mRNA sequence.  
 ACCESSION BU010277  
 VERSION BU010277.1 GI:17364159  
 KEYWORDS EST.  
 SOURCE Oryzias latipes (Japanese medaka)  
 ORGANISM Oryzias latipes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 Acanthomorpha; Acanthopterygii; Percomorph; Atherinomorpha;  
 Betoniformes; Adrianchthyidae; Oryziinae; Oryzias.  
 1 (bases 1 to 632)  
 Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.  
 Medaka EST Project in Takeda's lab  
 Unpublished  
 Contact: Tadasi Shin-I  
 JOURNAL  
 COMMENT Center For Genetic Resource Information  
 National Institute of Genetics  
 111 Yata, Mishima, Shizuoka 411-8540, Japan  
 Tel: 81-559-81-6856  
 Fax: 81-559-81-6855









source 1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
BASE COUNT 94 a 238 c 161 g 107 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.52e-12 Length: 600  
Score: 310.50 Matches: 73  
Percent Similarity: 58.39% Conservative: 7  
Best Local Similarity: 53.28% Mismatches: 35  
Query Match: 27.55% Indels: 23  
DB: 12 Gaps: 3

US-09-595-947E-10 (1-214) x BG808248 (1-600)

QY 91 ATGGLUARGAAsnArgMethHisAspLeuAAsnSerAlaLeuAspAlaLeuArgGlyValLeu 110  
DB 1 CCGGAGCGGACCGCATGACCAACCTAAACCGCGCTGACGCGCTGCGGAGGTCGCG 60  
QY 111 ProThrPheProAspAspAlaAlaLeuThrIleGluThrLeuArgPheAlaHisAsn 130  
DB 61 CCGACCTTCCCGAGGATCCCAAGCTCAGGAAGATGAGACGCTGCGCTTCCGCCACAT 120  
QY 131 TTTTLeuAlaLeuThrIleuArgIleAlaAspHisSerLeuThrAla----- 148  
DB 121 TACATCTGGGGCTCACCGAGACTCTGCGCTGGGACCACTGCGCGCGCGCTGTC 180  
QY 149 -----LeuGluProProAlaProHisCysGlyGlu 158  
DB 181 CTCGAGGGGGCGCTTTCACGAGAGCGGTCCTCTGAGCCCGGAGCTCGCTCGCGCC 240  
QY 159 LeuGlySerProGlyGlyProProGlyAspTTPGlySerLeuThrProValSerGln 178  
DB 241 AGCGGGGACAGCCCTTCTCACCTTCTCTGAGCTGACCAACAGCCCGCGCTCATCC 300  
QY 179 AlaGlySerLeuSerPro-----AlaAlaSerLeuGluGluArgPro 192  
DB 301 TCCAACTCCAGTCCCATATACAGCTGCACTTATGCGCCGCTAGGCC -CGGCTCAGAGCT 359  
QY 193 GlyLeuLeuGlyAlaThrSerSerAla-----CysLeuSerPro 205  
DB 360 GGAACACTGCGACGCCCACTCTCGGAGAGCATCGTTATGCGCTCACT 410

RESULT 12  
LOCUS BU924937 600 bp mRNA linear EST 30-OCT-2002  
DEFINITION 7103-91 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,  
RNA sequence.  
ACCESSION BU924937  
VERSION BU924937.1 GI:24428820  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 600)  
Mu.X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,  
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L., Klein,W.H.,  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
JOURNAL 21671825  
MEDLINE 11812828  
PUBMED  
COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646

FEATURES Fax: 713 790 0329.  
Location/Qualifiers  
source 1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"  
BASE COUNT 99 a 194 c 222 g 82 t 3 others  
ORIGIN

Alignment Scores:  
Pred. No.: 2.77e-11 Length: 600  
Score: 299.00 Matches: 88  
Percent Similarity: 52.15% Conservative: 21  
Best Local Similarity: 42.11% Mismatches: 32  
Query Match: 26.53% Indels: 68  
DB: 13 Gaps: 10

US-09-595-947E-10 (1-214) x BU924937 (1-600)

QY 26 GluAspGluVal-----ThrCysProThrSerAlaProProSerPro 39  
DB 39 GAAGAGAGGATCTGATGCTGCTGCGCTTCCCGGCTCGGACCTGACCCCG 98  
QY 40 -----ThrArgThrProGlyAsnCys--- 46  
DB 99 ATGCTCTCAGCGCGGACGAGAGGAGGACGAGAGCTGCCGCGGCGCTCGCGGCT 158  
QY 47 -----AlaGluAlaGluGlu-----Gly 52  
DB 159 GGGCAGCGTGGAGCGGAGACCGGAGGGGTGACAGGCACTCCGCTCGGCTCGCGG 218  
QY 53 GlyCysArgGlyAlaProArgIleuArgAla-----ArgArgGly 66  
DB 219 GGTTCGCG-----CCAGGGCGGCTGCTGGGCTGATGACAGAGTGCAGAGCTGCCG 272  
QY 67 GlyArgSerArg-----ProLysSerGluLeuAlaLeuSerLysGlnArg 81  
DB 273 TCGGCTCAGCGGCGCGCTCTCCGAGGTGCCAAGACGGGAGACGGTCCAGCGCATCAG 332  
QY 82 ArgSerArgArgLysLysAlaAspAspArgIleuArgAlaGlyMethHisAspLeuAsnSer 101  
DB 333 AAGACCCCGAGCTTCAGAACCAACGCGGACGACCACTGACCACTTAAACGCC 392  
QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121  
DB 393 GCGCTGAGACGGCTGCGGAGGTGCTGCCACCTTCCCGAGAGATGCCAAGCTCAGAG 452  
QY 122 IleGluThrLeu--ArgPheAlaHisAsnTyrIleTTPAlaLeuThrGlnThrLeuArgI 141  
DB 453 ATCGAGAGCTTGGCCCTTCCGCCCATTTACATTTGGGGCTCACGAGACTCGGCC 512  
QY 141 LeuAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 161  
DB 513 TGGCGGAC-----CACATC----- 526  
QY 161 erProGlyGlyProProGlyAspTTPGlySerLeuThrSerProValSerGlnAlaGly 181  
DB 527 -----GCCGGGCGCGGTGCTGCTCCAGAGGGGGCGCTTTC-----ACGAGNGCGTGC 572  
QY 181 erLeuSerProAlaAlaSerLeu 188  
DB 573 TNCCTAGCGCGGAGACTCGGCTC 595

RESULT 13  
LOCUS AL540071 1001 bp mRNA linear EST 31-MAY-2003  
DEFINITION AL540071 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
CS0DF035YD19 5-PRIME, mRNA sequence.  
ACCESSION AL540071  
VERSION AL540071.2 GI:31264632

KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1001)  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
JOURNAL Full-length cDNA libraries and normalization  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12869886.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
2626.r, Contact: Feng Liang Email: fliang@litech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID: CS0DF035CB10QPL  
Location/Qualifiers  
1. 1001  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF035YD19"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: Brain; Vector: pCMVSPORT 6, 1st strand cDNA  
was primed with a NotI-oligo (dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."  
BASE COUNT 190 a 322 c 342 g 121 t 26 others  
ORIGIN

Alignment Scores:  
Pred. No.: 4.9e-11 Length: 1001  
Score: 298.50 Matches: 81  
Percent Similarity: 48.50% Conservative: 16  
Best Local Similarity: 40.50% Mismatches: 40  
Query Match: 26.49% Indels: 64  
DB: 9 Gaps: 8

US-09-595-947E-10 (1-214) x AL540071 (1-1001)

QY 31 CysProthrsAlaProProSerProthrsArgThrProGlyAsnCyS----- 46  
DB 421 TGTCTGCTGGATCGGCTCTCCCGCTTGGCGGCGCTGACCCCGCTGTCATCCAGCGCG 480  
QY 47 -----AlaGluAlaGluGluGlyGlyCysArgGlyAlaProArglys 60  
DB 401 ACGAAGAAGAGAGAGAGCGGCGGCGCTCAAGCGGCGCGCTCGGCAAGCGCGGCGC 540  
QY 61 LeuArgAlaArgArgGlyGlyArgSer----- 69  
DB 541 TTA---NGCCGGCAANNGCGGCGCGGCGCTTCTCCGGGCTCCGAGAGGCGTCCGG 597  
QY 70 -----ArgProlysSerGluLeuAla 76  
DB 598 CCCGACGCGCTGCTGGTCTGTACATTCGCAACGCGCGCTTCCCGGCGCGCGGCGC 657  
QY 77 LeuSerLysGln-----ArgArgSerArgArglys 86  
DB 658 GCTCTCCGAGGCGCCAGACGCGCCGACGCTGACGCGCATCAAGAACCGGTGACCTG 717  
QY 87 LysAlaAsnAspArgGlyAlaArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106  
DB 718 AAGGCAACAACCGGAGGAGAAACCGCATGCAACACTCAACGCGCACTGACCGCTG 777  
QY 107 ArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeuArg 126  
DB 778 CCGGAGTGTCTCCCACTTCCCGAGAGCGCCAAAGCTCAACAAGATCGAGACGCTGCGC 837

QY 127 PheAlaHisAsnTyrlleTrrAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146  
DB 838 TTCCGCCCACTAATCTGGGAGCTCACAGACCTTCCGCTGCGGCT----- 888  
QY 147 TyraLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProPro 166  
DB 889 -----MAYTCCGGCGG-GCGGCGTCC---GCGGCGAATGCGC 920  
QY 167 GlyAspTrpGlySerLeuTyrsSerProValSerGlnAlaGlySerLeuSerProAlaAla 186  
DB 921 GGG-----GGGCTCTCTTCTC-----GAGGCMRTGTTKRTAGMSCCGGAGCG 962

RESULT 14  
BE780690 1039 bp mRNA linear EST 20-OCT-2000  
LOCUS 601469349P1 NIH\_MGC\_67 Homo sapiens cDNA clone IMAGE:3872172 5',  
DEFINITION mRNA sequence.  
ACCESSION BE780690.1 GI:10201888  
VERSION BE780690.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1039)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILMN at:  
http://image.llnl.gov  
Plate: LHM9626 row: b column: 13  
High quality sequence stop: 692.  
Location/Qualifiers  
1. 1039  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3872172"  
/tissue\_type="retinoblastoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 67"  
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.75 kb. Library constructed by Life  
Technologies."  
BASE COUNT 228 a 351 c 337 g 122 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 6.96e-11 Length: 1039  
Score: 296.50 Matches: 95  
Percent Similarity: 49.33% Conservative: 16  
Best Local Similarity: 42.22% Mismatches: 66  
Query Match: 26.31% Indels: 53  
DB: 10 Gaps: 6

US-09-595-947E-10 (1-214) x BE780690 (1-1039)

QY 4 GlnProSer-----GlyAlaProThrValGlnValThrArgGluThrGluArgSer 20  
DB 244 CAGCCCGCTTGAACCTCGCATCTCCGACTCGACTGCGCCAGACGAGCGGAGTACCC 303  
QY 21 PheProArgAlaSerGluAspGluValThrCys-----ProThr 33  
DB 304 TATCCGGTCTCTCAACC-GACGAGGAGAGACTGTGCGACACTCCAAACAGCAGCTCCGCT 362



---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:43:11 ; Search time 5442 Seconds

(without alignments)  
10975.387 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1 ggcagtagcagagagagcag.....agagtgcctcattcagtggt 1460

Sequence: 1 ggcagtagcagagagagcag.....agagtgcctcattcagtggt 1460

Scoring table: IDENTITY NUC  
Gapop 10-0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vit:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1458.4	99.9	1491	6 A91167	A91167 Sequence 1
2	1458.4	99.9	1491	6 BD023626	BD023626 Polypept
3	1458.4	99.9	1491	10 RNRELAXT	Y10619 R.norvegicus
4	1407	96.4	258815	2 AC127817	AC127817 Rattus no
5	932.2	68.0	215050	2 AC127417	AC127417 Mus muscu
6	976.6	66.9	138070	2 AC109783	AC109783 Mus muscu
7	900.6	61.7	1861	10 AF364300	AF364300 Mus muscu
8	819.8	56.2	5567	10 MMMA7H4B	Y09167 M.musculus
9	682.4	46.7	861	6 AX698801	AX698801 Sequence
10	682.4	46.7	861	10 MMU76208	U76208 Mus musculu
11	517.4	35.4	5340	9 AF234829	AF234829 Homo sapi
12	515.8	35.3	165110	9 AL450311	AL450311 Human DNA
13	512.6	35.1	173341	2 AC021954	AC021954 Homo sapi
14	456.8	31.3	1330	9 HSA133776	A133776 Homo sapi
15	210.6	14.4	170896	2 AC011010	AC011010 Homo sapi
16	147.6	10.1	1268	6 AR023709	AR023709 Sequence
17	147.6	10.1	1268	6 AR225842	AR225842 Sequence
18	147.6	10.1	1268	6 HSU63842	U63842 Human neuro
19	147.6	10.1	1675	9 BC008687	BC008687 Homo sapi
20	147.6	10.1	1717	9 BC028226	BC028226 Homo sapi
21	147.6	10.1	79970	2 AC011428	AC011428 Homo sapi
22	147.6	10.1	134506	2 AC005738	AC005738 Homo sapi
23	147.6	10.1	151044	2 AC011432	AC011432 Homo sapi
24	145.2	9.9	1527	6 AR308547	AR308547 Sequence
25	145.2	9.9	1527	10 RNU67777	U67777 Rattus norv
26	145.2	9.9	258118	2 AC112007	AC112007 Rattus no
27	145.2	9.9	258880	2 AC095430	AC095430 Rattus no
28	140.4	9.6	735	10 MMU67776	U67776 Mus musculu
29	140.4	9.6	738	6 AR308548	AR308548 Sequence
30	140.4	9.6	932	10 MMMA7H4C	Y09166 M.musculus
31	140.4	9.6	1315	10 MMU63841	U63841 Mus musculu
32	140.4	9.6	1333	6 AR023715	AR023715 Sequence
33	140.4	9.6	1333	6 AR225848	AR225848 Sequence
34	140.4	9.6	71538	2 AC118243	AC118243 Mus muscu
35	140.4	9.6	149268	2 AC124395	AC124395 Mus muscu
36	139.8	9.6	1385	10 MMU76207	U76207 Mus musculu
37	139.8	9.6	1412	10 MMMA7H4A	Y07621 M.musculus
38	139.8	9.6	1719	5 AB065284	AB065284 Cyrops py
39	139.8	9.6	10393	10 AF303001	AF303001 Mus muscu
40	139.8	9.6	123855	2 AC102600	AC102600 Mus muscu
41	139.2	9.5	770	5 AF123884	AF123884 Gallus ga
42	139.2	9.5	790	5 GGA012659	AV012659 Gallus ga
43	139.2	9.5	1880	5 AF303000	AF303000 Gallus ga
44	139.2	9.5	2370	9 BC036847	BC036847 Homo sapi
45	139.2	9.5	6123	9 AF303002	AF303002 Homo sapi

#### ALIGNMENTS

RESULT 1  
A91167  
LOCUS A91167 1491 bp DNA linear PAT 22-JAN-2000  
DEFINITION Sequence 1 from Patent WO9827206.  
ACCESSION A91167  
VERSION A91167.1 GI:6740202  
KEYWORDS  
ORGANISM Rattus sp.  
SOURCE Rattus sp.  
Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 1491)  
AUTHORS Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.  
JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;

Pred. No. is the number of results predicted by chance to have a





FEATURES	source	Location/Qualifiers
BASE COUNT	307 a 487 c 413 g 284 t	
ORIGIN		
Query Match	99.9%; Score 1458.4; DB 6; Length 1491;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 1459; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
PR	19-DEC-1996 FR 96/15651	
PI	CHRISTINE I CARD LIEPKAMP, JACQUES MALLET, PHILIPPE RAVASSARD, PC	
	C07K14/47, A61K31/711, A61K35/76, A61K38/00, A61K48/00, A61P25/00, PC	
	C12N15/09,	
PC	C12N15/00, A61K37/02	
CC	Strandedness: Single;	
CC	Topology: Linear;	
PH	Key Location/Qualifiers	
	1. 1491	
	/organism="Rattus sp."	
	/mol_type="genomic DNA"	
	/db_xref="taxon:10118"	
Query	1	GCAGGTACCGAGAGGAGCATGTCCTCTGGGCCCCCGTTGCTGATTGGCCCGTGCAACAGCA 60
Ds	1	GCAGGTACCGAGAGGAGCATGTCCTCTGGGCCCCCGTTGCTGATTGGCCCGTGCAACAGCA 60
Query	61	GCAGGCCGGGACGGGACGGCTCCGTGCTCCGGGAGACAGATTAAGGTGCAAGGGACACA 120
Ds	61	GCAGGCCGGGACGGGACGGCTCCGTGCTCCGGGAGACAGATTAAGGTGCAAGGGACACA 120
Query	121	CGATTAGAGCTCAGAAAGTCCCTCTGGGCTCTACCACTGCA CAGAGCCGAGAGACCCCT 180
Ds	121	CGATTAGAGCTCAGAAAGTCCCTCTGGGCTCTACCACTGCA CAGAGCCGAGAGACCCCT 180
Query	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTATCTCAAGCGAGGGCGCTGCACTCAG 240
Ds	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTATCTCAAGCGAGGGCGCTGCACTCAG 240
Query	241	CAAACTTCGAAGCGAGAGGGGTTACGCTATCCACCGCTGTTGACTTGCACCAACC 300
Ds	241	CAAACTTCGAAGCGAGAGGGGTTACGCTATCCACCGCTGTTGACTTGCACCAACC 300
Query	301	GCAGCTCTGTGTTCTTTGAGCCCGAGATTAAGTAACTTTAGAACTCTCAAAAGG 360
Ds	301	GCAGCTCTGTGTTCTTTGAGCCCGAGATTAAGTAACTTTAGAACTCTCAAAAGG 360
Query	361	TAGAAAGGGGAGTGGTGGGCTGCTCTAGTCCCGGTGAGTGA CTTAAGTCAAG 420
Ds	361	TAGAAAGGGGAGTGGTGGGCTGCTCTAGTCCCGGTGAGTGA CTTAAGTCAAG 420
Query	421	ACTGTCAACACCCCTCTCAATTTTTCCTCAACCTCAGATGAGCGCTCATCCTTGATG 480
Ds	421	ACTGTCAACACCCCTCTCAATTTTTCCTCAACCTCAGATGAGCGCTCATCCTTGATG 480
Query	481	CGCCACCATCCAAAGTGTCCAAAGAACCCACCAACCTTTCGGGAGCTCGACACAG 540
Ds	481	CGCCACCATCCAAAGTGTGTCCAAAGAACCCACCAACCTTTCGGGAGCTCGACACAG 540
Query	541	AAGTCTCAGTTCCAATTCACCCCACTTAGCCCACTCTCTGTA CCGAGGACCTGCTCCG 600
Ds	541	AAGTCTCAGTTCCAATTCACCCCACTTAGCCCACTCTCTGTA CCGAGGACCTGCTCCG 600
Query	601	AAAGCAAGACAGGTGA CTGCGAGGAGCATCGAAGAGCTCGTGCGCGCGCGAGAGG 660
Ds	601	AAAGCAAGACAGGTGA CTGCGAGGAGCATCGAAGAGCTCGTGCGCGCGCGAGAGG 660
Query	661	GGAAGAGGCGCAAGGAGGAGTGGACCTAGAGAGGAGGAGCGAGAGCGGCGCAAGAGG 720
Ds	661	GGAAGAGGCGCAAGGAGGAGTGGACCTAGAGAGGAGGAGCGAGAGCGGCGCAAGAGG 720
Query	721	CCAAAGACCGGAGGCGCAACCGCATGCA CAACCTTAACCTCGGCTGAGTGGCTGCGG 780
Ds	721	CCAAAGACCGGAGGCGCAACCGCATGCA CAACCTTAACCTCGGCTGAGTGGCTGCGG 780
Query	781	GTGTCTGCTCCACTTCCCGGATGACGCCAATTTAAGATTCAGAACCTTGCGCTTG 840

D	b		781	GTCGCTGGCCACCTTCCCGGATGAGCCAAACTTA	CMAAGATGAGAACCCTTGCTTCG	840
O	y		841	CCCAACAATCATTTGGGCACTGA	CTAAGAGCTGGCATAAGCCGACCAAGCTTCTAACG	900
D	b		841	CCCAACAATCATTTGGGCACTGA	CTAAGAGCTGGCATAAGCCGACCAAGCTTCTAACG	900
O	y		901	GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAACCCGGGAGGGGGGCTCCAGCGGCG		960
D	b		901	GGCCCCGAGCCCCCTGTGCCCTGTGGGGAGCTGGGAAACCCGGGAGGGGGGCTCCAGCGGCG		960
O	y		961	ACTGGGGCTCTATCTTACTCTCCCAAGTTTCCCAAGCTGGTAGGCTGAGCCCAAGCTTCAT		1020
D	b		961	ACTGGGGCTCTATCTTACTCTCCCAAGTTTCCCAAGCTGGTAGGCTGAGCCCAAGCTTCAT		1020
O	y		1021	TGGAGAGATTCCCTGGGCTTGACAGGTGCAGCTCCCATCTGTCTCTCCCGGAGCACCC		1080
D	b		1021	TGGAGAGATTCCCTGGGCTTGACAGGTGCAGCTCCCATCTGTCTCTCCCGGAGCACCC		1080
O	y		1081	TGGAGTTCTAGACTTCTGTGTGAAGGGGCCCAAACAGGCCCCTGGGCGGTGGGCGCTGGCAG		1140
D	b		1081	TGGAGTTCTAGACTTCTGTGTGAAGGGGCCCAAACAGGCCCCTGGGCGGTGGGCGCTGGCAG		1140
O	y		1141	AAAAGGAGGAGAGTCAAGACTGTCTGAAATAGGAAGTAGTAGGAGCACTCGACATCTCGC		1200
D	b		1141	AAAAGGAGGAGAGTCAAGACTGTCTGAAATAGGAAGTAGTAGGAGCACTCGACATCTCGC		1200
O	y		1201	CCCTTCTGGCTTTTCATTAGTCAGGTCCCTGTATTTAACCAAGATTGCGACAGTTCTTGCT		1260
D	b		1201	CCCTTCTGGCTTTTCATTAGTCAGGTCCCTGTATTTAACCAAGATTGCGACAGTTCTTGCT		1260
O	y		1261	GCTGTGCTGTCACAAAGATTTGCAAGAGCTGATCTCTCTTAAACCTTCCTACGTGTGGC		1320
D	b		1261	GCTGTGCTGTCACAAAGATTTGCAAGAGCTGATCTCTCTTAAACCTTCCTACGTGTGGC		1320
O	y		1321	ACCTCAAACTCCGCGCTCCAAGCAGAGGAGCCGTAGACACTAAATAGTTGGAGACTCC		1380
D	b		1321	ACCTCAAACTCCGCGCTCCAAGCAGAGGAGCCGTAGACACTAAATAGTTGGAGACTCC		1380
O	y		1381	ATACTTCTGTGTGACTCGCCCTCTTTCAATCTGCGGCGCTCCACACCGCTTTCTCC		1440
D	b		1381	ATACTTCTGTGTGACTCGCCCTCTTTCAATCTGCGGCGCTCCACACCGCTTTCTCC		1440
O	y		1441	AGAGTGACCTAATCCAGTGT	1460	
D	b		1441	AGAGTGACCTAATCCAGTGT	1460	

RESULT 3  
RNRRLAXT                      1491 bp       mRNA       linear       ROD 06-MAY-1997

LOCUS                          R.nrveglus mRNA for transcriptional regulator, Relax.

DEFINITION                   Y10619

ACCESSION                   Y10619.1 GI:2072737

VERSION                      Relax; transcriptional regulator.

KEYWORDS                    Rattus norvegicus (Norway rat)

SOURCE                      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE                    1 Ravassard, P., Chatall, F., Maillet, J. and Icard-Liepkhalns, C.  
AUTHORS                      Relax, a novel rat bHLH transcriptional regulator transiently expressed in the ventricular proliferating zone of the developing central nervous system  
J Neurosci. Res. 48 (2), 146-158 (1997)  
MEDLINE                      97276390  
PUBMED                      9130143  
REFERENCE                    2 (bases 1 to 1491)  
AUTHORS                      Ravassard, P.  
TITLE                         Direct Submission  
JOURNAL                      Submitted (20-JUN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,  
Hospital de la Pltie Salpetriere, 83 Bd. de l'Hopital, F-75013



Bryant, N., Bahay, C., Butch, P., Butrell, K., Calderon, E.,  
 Cardenas, V., Carter, K., Cavazos, I., Cesari, H., Center, A.,  
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
 Davila, M. L., Davis, C., Davy-Carrroll, L., De Ande, C., Dedrich, D.,  
 Delgado, O., Denison, S., Desmo, C., Ding, Y., Dim, H., Divya, K.,  
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duvall, B., Eaves, K.,  
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
 Gduregeyigis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
 Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,  
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,  
 Hollins, B., Howells, S., Huliy, S., Hune, J., Ididit, D., Jackson, A.,  
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
 Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J.,  
 Lorenshewa, L., Louiseged, H., Lozado, R. O., Lu, X., Ma, J.,  
 Mageswari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangum, A.,  
 Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
 Pasternack, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,  
 Plopper, F., Polidexter, A., Popovic, D., Primus, B., Pu, L.,  
 Pizzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
 Rella, B., Rella, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,  
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,  
 Shetty, J., Shivatsbeyan, A., Sisson, I., Sitter, C. D., Smaj, D.,  
 Sneed, A., Sodergren, E., Song, X.-Z., Sorrell, R., Sosa, J.,  
 Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,  
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trojes, Z., Umani, K.,  
 Valas, R., Vera, V., Villanana, D., Waliron, L., Walker, B., Wang, J.,  
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
 Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K.,  
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
 Wedelhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
 Weinstein, G., and Gibbs, R. A.  
 Direct Submission  
 2 (bases 1 to 258815)  
 Unpublished  
 Direct Submission  
 2 (bases 1 to 258815)  
 Morley, K. C.  
 Direct Submission  
 Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 258815)  
 Rat Genome Sequencing Consortium.  
 Direct Submission  
 Submitted (19-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 19, 2002 this sequence version replaced gi:23912578.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine

```

Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GZXS
Center clone name: CH230-259G16
-----
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228224; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.
*
* Location/Qualifiers
*     1..258815
*     /organism="Rattus norvegicus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:10116"
*     /clone="CH230-259G16"
*     1..1609
*     /note="wgs_end_extension
*     clone_end:5p6"
*     complement(2865..3659)
*     /note="Clone boundary
*     clone_end:5p6
*     site:_
*     end_sequence:BT169379"
*     224098..225157
*     /note="Clone boundary
*     clone_end:T7
*     site:_
*     end_sequence:BT169378"
*
BASE COUNT      59351 a 56888 c 58131 g 56821 t 27624 others
ORIGIN
Query Match      96.4%; Score 1407; DB 2; Length 258815;
Best Local Similarity 99.1%; Pred. No. 2,5e-305;
Matches 1446; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
QY      4  GGTAGCGAGAGGAGGAGTCCCTGGGCGCCCGCTGTGCTGATTGGCCGTGGACAGGACGCA 63
Db      247404  GGGCTCGCTGGCGCAGTCCCTGGGCGCCCGCTGTGCTGATTGGCCGTGGACAGGACGCA 247463
QY      64  GCGCGGACGACGCGCTCCCTGGTCCGGGAGAGCAGATTAAAGCGGCCAGAGGGG -ACACAG 122
Db      247464  GCGCGGACGACGCGCTCCCTGGTCCGGGAGAGCAGATTAAAGCGGCCAGAGGGGACACAG 247522
QY      123  ATTAGAGCTCAGAAGTCCCTGGGTCTCACACTGCACAGAGAGCGAGAGACCCCTCC 182
Db      247524  ATTAGAGCTCAGAATACTCCCTGGGTCTCACACTGCACAGAGCGAGAGACCCCTCC 247588
QY      183  GAGCTTCTTGTGCTGCTCCAGCGCAATTTCCTCCAGGCGAGGCGCCTGCAGTTCAGCA 242
Db      247584  GAGCTTCTTGTGCTGCTCCAGCGCAATTTCCTCCAGGCGAGGCGCCTGCAGTTCAGCA 247643
QY      243  AACTTCGAAGGACGACAGGGGTCACTATCCACCGCGCTTGACTCTGACC -ACCG 301

```

```

Db 247644 AAATTCGAAACGAGAGAGGGGTTGAGCTATCCAGCCGCTGCTTGAACCAACCCG 247703
Qy 302 CAGCTCTCTGTTCTTTTGAAGCCCGAGTAATAGTAATATTGAAGACTCCAAAGGT 361
Db 247704 CAGCTCTCTGTTCTTTTGAAGCCCGAG-MACTAGTAATATTGAAGCCCTCCAAAGGT 247762
Qy 362 AGAAGAGGGAGTGGTGGGCGTACTTAAGTCCCGCGTGAAGTACTTAAAGTCAAGA 421
Db 247763 AGAAGAGGGAGTGGTGGGCGTACTTAAGTCCCGCGTGAAGTACTTAAAGTCAAGA 247822
Qy 422 CTGTCAACACCCCTCTCATTTTTCCTCAACCTAGAGATGGCGCTCATCCCTTGATGC 481
Db 247823 CTGTCAACACCCCTCTCATTTTTCCTCAACCTAGAGATGGCGCTCATCCCTTGATGC 247882
Qy 482 GCCCACCATCCAAATGTCTCCAAAGAGACCAGCAACCTTTTCCGAGCCTTGACCAAGA 541
Db 247883 GCCCACCATCCAAATGTCTCCAAAGAGACCAGCAACCTTTTCCGAGCCTTGACCAAGA 247942
Qy 542 AGTGTGAGTTCAATTCACACCCCACTTACCTGATCCGAGGAGCTGCTCGGA 601
Db 247943 AGTGTGAGTTCAATTCACACCCCACTTACCTGATCCGAGGAGCTGCTCGGA 248002
Qy 602 AGCAGAAAGAGTACTGCTCCGAGGAGCATCGAGAGTCTGCTGCGGCGCGGAGGCG 661
Db 248003 AGCAGAAAGAGTACTGCTCCGAGGAGCATCGAGAGTCTGCTGCGGCGCGGAGGCG 248062
Qy 662 CAACAGGCCCAAGAGCGAGTTGGCACTGAGCAACAGAGCAGAGCGGCGGAGAAAGGC 721
Db 248063 CAACAGGCCCAAGAGCGAGTTGGCACTGAGCAACAGAGCAGAGCGGCGGAGAAAGGC 248122
Qy 722 CAACAGCGGAGGCGCAACCGCATGCAACCTTAACTCCGCGTGAATGCGCTGCGCG 781
Db 248123 CAACAGCGGAGGCGCAACCGCATGCAACCTTAACTCCGCGTGAATGCGCTGCGCG 248182
Qy 782 TGTCTCTGCTGCTTCTCCGAGTGAAGCCCAACTTAAGATGAGAACCTTGCCTTGC 841
Db 248183 TGTCTCTGCTGCTTCTCCGAGTGAAGCCCAACTTAAGATGAGAACCTTGCCTTGC 248242
Qy 842 CCACAACTAATTTGGGAGCTGAGCTGAGACGCTGCGATAGCGAGCAACAGCTTCTACG 901
Db 248243 CCACAACTAATTTGGGAGCTGAGCTGAGACGCTGCGATAGCGAGCAACAGCTTCTACG 248302
Qy 902 CCCCAGACCCCTGCTGCTGCGAGCTGAGAAACCCGAGAGGCGGCTTCAAGCGCGA 961
Db 248303 CCCCAGACCCCTGCTGCTGCGAGCTGAGAAACCCGAGAGGCGGCTTCAAGCGCGA 248362
Qy 962 CTGGGGCTCTATCTACTTCCCAAGTTTCCCAAGCTGTAAGCTGACCCCAAGCTTCA 1021
Db 248363 CTGGGGCTCTATCTACTTCCCAAGTTTCCCAAGCTGTAAGCTGACCCCAAGCTTCA 248422
Qy 1022 GGAGAGATTTCCCTGGCGTGAAGTGGCCAGCTCCCATCTGTCMGTCCCGGAGACCT 1081
Db 248423 GGAGAGATTTCCCTGGCGTGAAGTGGCCAGCTCCCATCTGTCMGTCCCGGAGACCT 248482
Qy 1082 GGTGTTCTCAGACTTCTTGTGAAGGCGCCAAACAGGCGCTGGGCGTGGGCGTGGAGA 1141
Db 248483 GGTGTTCTCAGACTTCTTGTGAAGGCGCCAAACAGGCGCTGGGCGTGGGCGTGGAGA 248542
Qy 1142 AAGGAGGAGTGAAGCTGTTGAATGAAGTGAAGGCACTTGACACTTCTGCGC 1201
Db 248543 AAGGAGGAGTGAAGCTGTTGAATGAAGTGAAGGCACTTGACACTTCTGCGC 248602
Qy 1202 CTTTCTGCTTCTTCTTGAAGTGGTCCGATTTAAACAGAGATTGGACAGTTTCTTGCTG 1261
Db 248603 CTTTCTGCTTCTTCTTGAAGTGGTCCGATTTAAACAGAGATTGGACAGTTTCTTGCTG 248662
Qy 1262 CTGTGCGTGAACAAGATATTGAGAGTGAATCTCTTAAACCTCTCTCAGTGGGCA 1321
Db 248663 CTGTGCGTGAACAAGATATTGAGAGTGAATCTCTTAAACCTCTCTCAGTGGGCA 248722
Qy 1322 CTTCAAACTCCGCTCCAGAGAGAGAGAGCTGTAAGCACTAATATGTTGGAGACTCCCA 1381
Db 248723 CTTCAAACTCCGCTCCAGAGAGAGAGAGCTGTAAGCACTAATATGTTGGAGACTCCCA 248782

```

```

Qy 1382 TACTTCTGCTGAGCTCGCCCTTTTCAATCTGCGGCGCTCCAGCAGCTTTTCTCA 1441
Db 248783 TACTTCTGCTGAGCTCGCCCTTTTCAATCTGCGGCGCTCCAGCAGCTTTTCTCA 248842
Qy 1442 GAGTGAACCTAATCCAGTGT 1460
Db 248843 GAGTGAACCTAATCCAGTGT 248861

RESULT 5
AC127417
LOCUS
DEFINITION
AC127417
AC127417
AC127417.2 GI:24137619
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE
Mus musculus
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Oct 19, 2002 this sequence version replaced gi:21759524.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: M.BA0459M02
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.99019
Consensus quality: 212279 bases at least Q40
Consensus quality: 213137 bases at least Q30
Consensus quality: 213590 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 217944; sum-of-ctnigs
Quality coverage: 11.89 in Q20 bases; agarose-fp
Quality coverage: 10.54 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4227: contig of 4227 bp in length
* 4228 4327: gap of unknown length
* 4328 16589: contig of 12262 bp in length
* 16590 16689: gap of unknown length
* 16690 27318: contig of 10629 bp in length

```

```

* 27319 27418: gap of unknown length.
* 27419 49797: contig of 22379 bp in length
* 49798 49897: gap of unknown length
* 49898 79378: contig of 29481 bp in length
* 79379 79478: gap of unknown length
* 79479 132595: contig of 53117 bp in length
* 132596 132695: gap of unknown length
* 132696 215050: contig of 82355 bp in length.
FEATURES
source 1..215050
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-459M2"
misc_feature 1..4227
/note="assembly_name:Contig31"
misc_feature 4328..16589
/note="assembly_name:Contig32"
misc_feature 16690..27318
/note="assembly_name:Contig33"
misc_feature 27419..49797
/note="assembly_name:Contig34"
misc_feature 49898..79378
/note="assembly_name:Contig35"
misc_feature 79479..132595
/note="assembly_name:Contig36"
misc_feature 132696..215050
/note="assembly_name:Contig37"
BASE COUNT 56024 a 52513 c 52885 g 53012 t 616 others
ORIGIN
Query Match 68.0% Score 992.2; DB 2; Length 215050;
Best Local Similarity 83.9% Pred. No. 2.8e-212;
Matches 1275; Conservative 0; Mismatches 178; Indels 66; Gaps 11;
QY 4 GGTAGCAGAGAGACAGTCCCTGGGCCCCCGCTGCTGATTTGGCCCGTGGCAGAGCA 63
Db 205068 GGGCTCGGTGGCCAGCCCTGGGCCCCCGCTGATTTGGCCCGTGGTCAAGAGCA 205127
QY 64 GCCCGGAGGAGCGCTCTGTGTC-CGGGAGAGCAGATTAAGCGTGGCAGGAGACACG 122
Db 205128 GCCCGGAGGAGCGCTCTGTGTCGGGGGAGAGAGAGATTAAGCGTGGCAGGAGACACG 205187
QY 123 ATT-----AGCAGCTCAGAGTCCCTGGGCTCAGCAGTGA-CAGAGGCGAGAGACC 177
Db 205188 ACTTGACATGACGCTCAGAAATCCCTGGGCTCAGCAGTGAAGTGGAGTACT 205247
QY 178 CTTCCAGCTCTTGTGCTGCTCCAGACGCAATTACTCCAGGAGGAGCGCTGACACT 237
Db 205248 CTTCCAGCTCTTGTGCTGCTCCAGACGCAATTACTCCAGGAGGAGCGCTGACACT 205307
QY 238 CAGCAAACTTGAAGCAGAGAGGGTTCACTATCCACCGCTGCTTGACTTGACC- 296
Db 205308 TAGCAGAACTTCAAGGAGAGAGAGGCTCAGCTATCCACTGCTTGAACATGACC 205367
QY 297 -----ACCGCAGCTCTGCTCTTTTGAACCGGAGTA 330
Db 205368 TATCCACTGCTGCTTGTCACTGACTGACTGCTGCTCTTCTTCTTGAAGTGGAG-A 205426
QY 331 ACTAGTAACTTTAGAACCTCCAAAGGTAGAGAGGGAGTG-----GGTGG 380
Db 205427 ACTAGTAACTTTAGAACCTCCAAAGGTAGAGAGGGAGGGGGGGGTGTGTGGG 205486
QY 381 GCGTACTTACTCCCGGTGAGTGAAGCTTAACTCAGAGACTGTCAACCC--CCTTC 438
Db 205487 GGATATCTGTGCTCCCGGTGAGTGAAGCTTAACTCAGAGAGGCTGGACACACACACTTC 205546
QY 439 CATTTTCCCAACTCAGATGAGAGGCTCATCCCTTGGAGCGCCCAACCAATCCAAAGT 498
Db 205547 CATTTTCCCAACTCAGATGAGAGGCTCATCCCTTGGAGCGCTCAACCAATCCAAAGT 205606
QY 499 CCAAGAGACCAGCAACCTTCCCGGAGCTCGAGACCAAGAGTGTCAATT 558

```

```

Db 205607 CCCAGAGACACAAACCTTTCCCGAGCTCGACCAAGTGTCAATT 205666
QY 559 CCAACCCAGTACCCCACTCTGTACCGAGGAGCTGTCCGAAGAGAGAGTACT 618
Db 205667 CCAACCCAGTACCCCACTCTGTACCGAGGAGCTGTCCGAAGAGAGTACT 205726
QY 619 GCCAGAGGACATGAGAGAACCTCTGTGCGGCGCGAGGAGCGCAACAGGCCAAGAGCG 678
Db 205727 GCCAGAGGACATGAGAGAACCTCTGTGCGGCGCGAGGAGCGCAACAGGCCAAGAGCG 205786
QY 679 AGTTGGCACTGAGACAGACCGCAAGAGCGGCGCAAGAGGCCCAACGACCGGAGCGCA 738
Db 205787 AGTTGGCACTGAGACAGACCGCAAGAGCGGCGCAAGAGGCCCAACGATCGGAGCGCA 205846
QY 739 ACCGATGACACAACTTAACTCCGCGGTGATGGCGTGGCGGGTCTGCGCACTTCC 798
Db 205847 ATCGATGACACAACTTAACTCCGCGGTGATGGCGTGGCGGGTCTGCGCACTTCC 205906
QY 799 CGGATGACGCAAACTTACAAAGATCGAGACCTGCGCTTCCGCAACATCAATTGGG 858
Db 205907 CGGATGACGCAAACTTACAAAGATCGAGACCTGCGCTTCCGCAACATCAATTGGG 205966
QY 859 CACTGACTCAGACGCTGCGCATAGCGGACCAAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
Db 205967 CACTGACTCAGACGCTGCGCATAGCGGACCAAGCTTCTACGGCCCCGAGCCCCCTGTGC 206026
QY 919 CCGTGGGGAGCTGGGAGAGCCCGGAGGGGGCTCCAGCGCGCATGCGGGCTTACTACT 978
Db 206027 CCGTGGAGAGCTGGGAGAGCCCGGAGGGTGGCTTCAAGGGAGTGGGGCTTACTACT 206086
QY 979 CCCAGATTCCCAAGCTGAGAGCTGAGCCCAAGCGCTTATGAGAGATTCCTGGCC 1038
Db 206087 CCCAGATTCCCAAGCGGGATTAAGTGAAGCCCAAGCGCTTATGAGAGATTCCTGGCC 206146
QY 1039 TGAAGTGCCAGCTCCCATCTGTCTGCTCCGAGCACCTGGTGTCTCAGACTTCT 1098
Db 206147 TGAAGTGCCAGCTCCCATCTGTCTGCTCCGAGGAGCATGTTGTCTCAGACTTCT 206206
QY 1099 TGTGAAGGGCCCAAGAGCGCTGGGCGGTGGCGCTGCGCAAGAAAGGAGAGTCAAG 1158
Db 206207 TGTGAAGAGCTGTCTGCTGTGGGTGGGTGTGAGTGAAGGAGAGGAGACAGAG 206266
QY 1159 CTGCTGAATGGAAGTATGAGAGCACTGAGACATCGCCCTTCTGGCTTCAATTA 1218
Db 206267 CTGCTGAATGGAAGTATGAGAGCACTGAGACATCGCCCTTCTGGCTTCAATTA 206326
QY 1219 GTAGGTCCCT-----GATTTAACAGAAATTCGACAGTTCTTGCTGCTGT 1265
Db 206327 CTGGAATCCCTAGCCCTTCACAGGGCTTAAGGCTTCTCATGGGTAC--CCTTGCTG 206384
QY 1266 GCGTGCACAAAGATATGAGAGCTGATCTCTTAAACCTCTCTCAGTGTGGCACCTC 1325
Db 206385 CTGCGCACACAAACATTTGGGGGCTGCTTCTTAAACCTCTCTGCTGACAGCAATC 206444
QY 1326 AAATCCCGCTCCAAAGAGAGAGCGCTAGACATTAATAGTTGGAGACTCCATACT 1385
Db 206445 AAATCTCGCTTCAAGATTTGGAATGTGATGACATCTAGTT--GGAGCTCCCATACT 206503
QY 1386 TCTGTGATCTCGCCCTCTTCAAAATCTGGGGCT-----CCAAACACGCTTTCGA 1441
Db 206504 TCTGTGATCTCGCCCTCTTCAAAATCTGGGGCTCCGAGCATCATCACTTTTCCA 206563
QY 1442 GAGTGAATTAATCAAGT 1460
Db 206564 GGGTGAATTAATCAAGT 206582

```

RESULT 6  
 AC109783/c AC109783 138070 bp DNA linear HTG 07-FEB-2002  
 LOCUS  
 DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17  
 unordered pieces.

ACCESSION	GI:18581594
VERSION	HTG_HTG5_PHASE1; HTG5_DRAFT.
KEYWORDS	Mus musculus (house mouse)
SOURCE	Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	
AUTHORS	Zutavern, T., Baatle, M.R., Baker, J.P., Ballja, V., Dedhia, N.N., de la Baetle, M., Katzenberger, F., Kuit, K., King, L., Kirchoff, K.A., Miller, B., Muller, S., Nascimento, L.U., O'Shaughnessy, A.L., Preston, R.R., Santos, L., Spiegel, L.A., Palmer, L., Yang, C., and Zutavern, T.
TITLE	Mouse Genomic Sequence
JOURNAL	Unpublished
REFERENCES	2 (bases 1 to 138070)
AUTHORS	McCombie, W.R.
TITLE	Direct Submission
JOURNAL	Submitted (07-FEB-2002) Litca Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
COMMENT	-----Genome Center ----- Center: Litca Annenberg Hazen Genome Center, Cold Spring Harbor Laboratory Center code: CSHL. Web site: http://www.cshl.org/genseq Contact: mcombie@cshl.org -----Project Information Project name: RP23-121F10 Clone name: RP23-121F10 Insert size: 173000; agarose-fp Insert size: 141616; sum-of-contigs Quality coverage: 4.00 in Q20 bases; agarose-fp Quality coverage: 3.70 in Q20 bases; sum-of-contigs -----
* NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
* 1 17785: contig of 17785 bp in length	
* 17786 17874: gap of unknown length	
* 17875 30356: contig of 12482 bp in length	
* 30357 30444: gap of unknown length	
* 30445 42306: contig of 11862 bp in length	
* 42307 42394: gap of unknown length	
* 42395 53598: contig of 11204 bp in length	
* 53599 53686: gap of unknown length	
* 53687 64435: contig of 10669 bp in length	
* 64356 64443: gap of unknown length	
* 64444 74016: contig of 9573 bp in length	
* 74017 74105: gap of unknown length	
* 74105 83366: contig of 9262 bp in length	
* 83367 83454: gap of unknown length	
* 83455 92355: contig of 8901 bp in length	
* 92356 92443: gap of unknown length	
* 92444 100821: contig of 8378 bp in length	
* 100822 100909: gap of unknown length	
* 100910 107529: contig of 6620 bp in length	
* 107530 107617: gap of unknown length	
* 107618 114066: contig of 6449 bp in length	
* 114067 114154: gap of unknown length	
* 114155 118873: contig of 4719 bp in length	
* 118874 118961: gap of unknown length	
* 118962 123619: contig of 4658 bp in length	
* 123620 123707: gap of unknown length	
* 123708 128240: contig of 4533 bp in length	
* 128241 128328: gap of unknown length	
* 128329 128382: contig of 4354 bp in length	
* 132683 132770: gap of unknown length	

FEATURES					
	source				
*	132771	136341:	contig of 3571 bp in length		
*	136342	136429:	gap of unknown length		
*	136430	138070:	contig of 1641 bp in length.		
			Location/Qualifiers		
			1..138070		
			/organism="Mus musculus"		
			/mol_type="genomic DNA"		
			/db_xref="taxon:10090"		
			/clone="RP23-121P10"		
BASE COUNT					
	35731 a	33657 c	33303 g	33954 t	1425 others
ORIGIN					
Query Match 66.9%; Score 976.6; DB 2; Length 138070;					
Best Local Similarity 83.5%; Pred. No. 9,2e-209;					
Matches 1263; Conservative 0; Mismatches 184; Indels 66; Gaps 11;					
QY	10	GAGAGGACAGTCCCTGGGCCCCCCTGTGATGTCGCCTTGGCAGCAGCACGCCCG	68		
Db	111713	GCATGGCCCAAGCCCTTAGGGCCCCCTCCGCTGATTGGCCCGTGAGGAGCAGCACGCCG	111654		
QY	69	GCAGGCAAGCTTCCTGTGTCGGGCAAGCAGATTAAGAAGCTGCCAGGGAGACAAGAT---	125		
Db	111653	CCAGGCAAGCTTCCTGGCGAAGCAGACAGATAAAAGCGGCCAGGAGAACAGATTTGC	111594		
QY	126	- AGCAGCTCAGAAAGTCCCTCTGGGCTCAACAATGA - CAGAGGCGAGGACCCTCTCG	183		
Db	111593	ATGCACTCAGAAATCCCTCTGGGCTCATCTGACAGAGTGTGATGATCCTCTCGG	111534		
QY	184	AGCTTCCTTGTGCTCCAGACGCAATTTACTCAGGCGAGGGCGCTCAGCTCAGCAA	243		
Db	111533	AGCTTTTGAAGACTTCCAGACGCAATTTACTCAGGCGAGGGCGCTCAGTTTAGCAG	111474		
QY	244	AACTTCGAAGCGACAGAGGGGTTTACGCTATCACCGCTGCTTGACTTGACC-----	296		
Db	111473	AACCTTCGAAGCGACAGAGGGGTTTACGCTATCACCGCTGCTTGACTTGATCCA	111414		
QY	297	-----ACCGCAGCTCTGTTCTTTTGGCCGCGAGTAATTAGG	336		
Db	111413	CTGCTCTTGTTCAGTGAAGCTTGTCTCTTATTTTGTAGTGGGAG-AACTAGG	111355		
QY	337	TAACTTTAGAACTCCAAAGGTAGAAAGGAGAGT-----GGTGGCGTAC	386		
Db	111354	TAACTTTAGAACTCCAAAGGTAGAAAGGTAGAAAGGTAGTGTGGGGGATAC	111295		
QY	387	TCTAGTCCCGCGTGAATGCTCTPAAGTCAAGACTGTCAACCCC--CTTCCATTTT	444		
Db	111294	TCTGTGTCCTCCGTGAATGCTCTPAAGTCAAGGTGTGCAACACACTTCCATTTT	111235		
QY	445	TTCCCAACTCAGAAAGGCGGCTCAATCCCTTGGATGCGCCACATCAAGTGTCCCAAG	504		
Db	111234	TTCCCAACTCAGAAAGGCGGCTCAATCCCTTGGATGCGCTCAATCAAGTGTCCCAAG	111175		
QY	505	AGACCAGCAACCTTTTCCCGAGGCTCGAGCACAGATGCTCAATTCCATTTCAACC	564		
Db	111174	AGACCAACCAACTTTTCCCGAGGCTCGAGCACAGATGCTCAATTCCATTTCAACC	111115		
QY	565	CACCTAGCCCACTCTCGTACCGAGGACTGCTCCAGCAAGAGCAGTGAATGCCGAG	624		
Db	111114	CACCTAGCCCACTCTCATACTTGAAGACTGCTCCAGCAAGAGTGAATGCCGAG	111055		
QY	625	GGACATTCAGAAAGTCCGTGGCGCGCGCGGAGGGGCAACAGGCCCAAGAGGAGTTGG	684		
Db	111054	GGACCTCGAGAAAGTCCCGCGCGCGCAAGGAGGGGCAACAGGCCCAAGAGGAGTTGG	110995		
QY	685	CACCTAGCAAGAGGAGCGAAGCCCGCGCAAGAGGCAACGACCGGAGCGCAACGCA	744		
Db	110994	CACCTAGCAAGAGGAGCGAAGCCCGCGCAAGAGGCAACGATGATCCGAGCGCAATCGCA	110933		
QY	745	TGCACAACCTTAATCTCGCGTGAATGCGCTGCGCGGTGCTCTGCCACCTTCCGGATG	804		
Db	110934	TGCACAACCTTAATCTGCGGTGAATGCGCTGCGCGGTGCTCTGCCACCTTCCGGATG	110875		
QY	805	ACGCCAACTTACAAAGATGAGACCTTCGCTTCCGCCCAACAATATTTGGGCACTGA	864		



```

Db 110874 ACGCCAACTTACAAAGATCAGACCTGCGCTTCCGCCACAACTACATCTGGGACATGA 110815
Oy 865 CTGACAGCTGCGCATAGAGGACACACAGCTTCTACGCCCCGAGGCCCTGAGCCCTGTG 924
Db 110814 CTCACACCTGCGCATAGAGGACACACAGCTTCTACGCCCCGAGGCCCTGAGCCCTGTG 110755
Oy 925 GAGAGCTGGGAAGCCCGGAGGAGGAGCTCAGCGGCGAGCTGGGGCTCTATCTACTCCAG 984
Db 110754 GAGAGCTGGGAGGCCCGGAGGAGGAGCTCAGCGGCGAGCTGGGGCTCTATCTACTCCAG 110695
Oy 985 TTTCCCAAGCTGTGAGCTGAGCCCAAGCTCTATTGAGAGAGTTCCCTGGCCCTGACAG 1044
Db 110694 TCTCCCAAGCTGTGAGCTGAGCCCAAGCTCTATTGAGAGAGTTCCCTGGCCCTGACAG 110635
Oy 1045 TGCCAGCTCCCGCATCTGTCTGCTCCGCGGACCCCTGCTGTCTCAGACTTCTGTGAA 1104
Db 110634 TGCCAGCTCCCGCATCTGTCTGCTCCGCGGACCCCTGCTGTCTCAGACTTCTGTGAA 110575
Oy 1105 GGGCCCAAAACAGGCGCTGCGGAGGAGGAGCTGCGGAGGAGGAGGAGCTGCTCT 1164
Db 110574 GAGAGCTGTGAGCTCTGGGTGTGGGTGTAGTGAAGGAGGAGGAGGAGGAGGAGCTCT 110515
Oy 1165 GAATGGAAGTGTGAGGAGCACTGAGCATCTGCCCCCTTCTGCTTCTTCACTTACTGAG 1224
Db 110514 GAGAGTGGAGGTGTGAGGAGCTCTCAGCATCTGCTTCTGCTTCTTCACTTACTGAG 110455
Oy 1225 TCCCT-----GATTAAACAGAGTTCGCAAGTTCTTCTGCTGCTGAGCCGTC 1271
Db 110454 TCCCTAGCCCTCTCAAGAGGCTTAACTAGGCTTCTCACTGAGGAGGAGGAGGAGGAGGAG 110397
Oy 1272 ACAAGGATATTTGAGGCTGATCTCTCTTAACTCTCTCAGTGTGGGACACTCAATC 1331
Db 110396 ACAGAGACATTTGGGGGCTGCTCTTCTTAACTCTCTCAGTGTGGGACACTCAATC 110337
Oy 1332 CCGCTCAAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1391
Db 110336 TCGCTCAAGAGATTTGAGAGTGTGAGCACTACCTAGTT--GAGAGCTCCATATCTTCCG 110278
Oy 1392 TGACTCGGCGCTTTTAAATCTGGGGCT-----CGAACACGCTTTCGAGAGGA 1447
Db 110277 TGAGTCTGCTCATTTAAATCTGCGGAGCTCCAGCAATCATCATCTTTTTCGAGGAGA 110218
Oy 1448 CCTATCCAGTGT 1460
Db 110217 CCTATCCAGTGT 110205

```

RESULT 7  
 AF364300 1861 bp DNA linear ROD 03-MAY-2001  
 LOCUS AF364300  
 DEFINITION Mus musculus neurogenin 3 gene, complete cds.  
 ACCESSION AF364300.1 GI:13937128  
 VERSION AF364300.1  
 KEYWORDS  
 SOURCE  
 ORGANISM Mus musculus (house mouse)

REFERENCE  
 AUTHORS Lee,J., Smith,S., Matada,H., Lin,J., Scheel,D., Wang,J., Mimitra,R. and German,M.  
 TITLE Regulation of the pancreatic pro-endocrine gene neurogenin 3  
 JOURNAL Diabetes (2001) In press  
 REFERENCE  
 AUTHORS Schwitzgebel,V. and German,M.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave, HSW1090, San Francisco, CA 94145-0534, USA

FEATURES  
 source 1..1861  
 /organism="Mus musculus"

```

Query Match 61.7%; Score 900.6; DB 10; Length 1861;
Best Local Similarity 86.2%; Pred. No. 1.7e-191;
Matches 1094; Conservative 0; Mismatches 129; Indels 46; Gaps 7;

CDS
mRNA
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1093..>1737
/product="neurogenin 3"
1093..1737
/codon_start=1
/product="neurogenin 3"
/protein_id="AAK5058.1"
/db_xref="GI:13937129"
/translation="MAPHPDALTIQVSPETPOEPFGASPDHEVLSNSTPSPPLIR
DCSRYVDRGSRKLRARGRNRKSTALSKORSRKRKANDERRRNNINSA
LDLRGVLPFPDDAKITKTEIRFPHNNTIMALTQTRIDHSHYGPFPVPCGLS
PFGGSNDWDSIVSPVSAQNLSPVSLSEFPGIQLVSSPSYLLPVLVSDFL"
BASE COUNT 397 a 560 c 537 g 367 t
ORIGIN
4 GGTAGCAGAGAGAGACAGTCCCTGGGAGCCCGCTTCTATTGGCCCGTGACACAGACAGA 63
594 GGGCTCGGTGGCGACAGCCCTGGGAGCCCTCCGCTAATTTGGCCCGTGAGCAGACAGA 653
64 GCCCGAGAGGACCGCTCTGATC--CGGAGAGAGATTAAGCGTGCAGAGGAGACACAG 122
654 GCCCGAGAGGACCGCTCTGGCCGGGGGAGAGAGATTAAGCGTGCAGAGGAGACACAG 713
123 ATT---AGAGCTCAGAGTCCCTCTGGGTCTCAACATGCA--CAGAGCCGAGAACCC 177
714 ACTTGACATGACAGCTCAAAAATCCCTCTGGGTCTCATCATGCACTGACAGTGTGAGTACT 773
178 CCTCGAGCTTCTTTGTGCTGCTCCAGAGCAATTTACTCCAGAGGAGGAGGAGGAGGAGGAG 237
774 CTTGAGAGCTTTTCTACAGCTTCCAGAGCAATTTACTCCAGAGGAGGAGGAGGAGGAGGAG 833
238 CAGCAAACTTTCAGAGAGAGAGAGGAGGAGTTCAGCTATCCACGCTGCTTGAACCTGAC 296
834 TAGCAGAACTTTCAGAGAGAGAGAGAGGAGGAGTTCAGCTATCCACGCTGCTTGAACCTG 893
297 -----ACCGAGCTCTGTTT-----ACCGAGCTCTGTTT-----ACCGAGCTCTGTTT 330
894 TATCAGCTGCTGCTTTCATCTGACATGACCTGCTGCTCTCTAATCTTTTGAAGTCGGAG-A 952
331 ACTAGTAACATTAGAACCTCCAAAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 380
953 ACTAGTAACATTAGAACCTCCAAAGGAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1012
381 GCGTACTCTAGTCCCGGTGAGTGAAGTCTTAACTGACAGACTG--TCACACCCCTTTC 438
1013 GGAATCTGTGCTCCCGGTGAGTGAAGTCTTAACTGACAGAGGCTGAGACACACACACTTC 1072
439 CATTTTTCCTCACTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 498
1073 CATTTTTCCTCACTCAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1132
499 CCCAAGAGCCAGAACCTTTTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 558
1133 CCCAAGAGCCAGAACCTTTTCCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1192
559 CCAACCCACCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 618
1193 CCAACCCACCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1252
619 GCCGAGGAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 678
1253 GCCGAGGAGCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1312
679 AGTTGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 738
1313 AGTTGACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1372
739 ACCGATGACAACTTAATCTCCGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 798

```

```

Db      1373 ATCGCATGACAGAACTGCACTGAGCGCTGAGATGCGCTGGCGGTCTCTGCGCACTTCC 1432
Qy      799 CGAGTGAAGCCCAATTTACAAAGATGAGACCCCTGCTTGGCCCAACAATTCATTTGGG 858
Db      1433 CGAGTGAAGCCCAATTTACAAAGATGAGACCCCTGCTTGGCCCAACAATTCATTTGGG 1492
Qy      859 CACTGACTCAGACGCTGCGCATAGCGGACCAACAGCTTCTACGCGCCCGAGCCCTGTGCG 918
Db      1493 CACTGACTCAGACGCTGCGCATAGCGGACCAACAGCTTCTACGCGCCCGAGCCCTGTGCG 1552
Qy      919 CCTGTGGGAGCTGGGAAAGCCCGGAGGAGGCTTCAGCGGCGACTGGGGCTTATCTACT 978
Db      1553 CCTGTGGGAGCTGGGAGGAGCCCGGAGGAGGCTTCAGCGGCGACTGGGGCTTATCTACT 1612
Qy      979 CCCCACTTTCCCAAGCTGTGAGCTGAGCGCCCAACAGCTCTATTGAGAGAGTTCCCTGGCC 1038
Db      1613 CCCCACTTTCCCAAGCTGTGAGCTGAGCGCCCAACAGCTCTATTGAGAGAGTTCCCTGGCC 1672
Qy      1039 TGCAGAGTGCAGCTCCCATCTGCTGCTGCTCCCGGAGCACTGTGTTCTCAGACTTCT 1098
Db      1673 TGCAGAGTGCAGCTCCCATCTGCTGCTGCTCCCGGAGCACTGTGTTCTCAGACTTCT 1732
Qy      1099 TGTGAAGAGGCCCAACAGGCGCTGGGCGCTGGGCGCTGGCAGAGAGGAGAGTCAAGG 1158
Db      1733 TGTGAAGAGAGCTGCTGCTGCTGCTGGGTTGGTGTGAGTGAAGAGGAGGAGCAAGAG 1792
Qy      1159 CTGTCTGAAGTGAAGAGTGTGAGGAGCACTCAGAGATCTCGCCCTTCTGGCTTTCATTA 1218
Db      1793 CCGTCTGAAGTGAAGAGTGTGAGGAGCTCTCAAGATCTCGCCCTTCTGGCTTTCATTA 1852
Qy      1219 GTCAGATCC 1227
Db      1853 CTGGATCC 1861

RESULT 8
MMATH4B
LOCUS   MMATH4B 5567 bp DNA linear ROD 25-OCT-2000
DEFINITION MMATH4B gene.
ACCESSION Y09167
VERSION Y09167.2 GI:11065737
KEYWORDS bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Cau,E., Gradwohl,G., Fode,C. and Guillemot,F.
TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron
JOURNAL Development 124 (8), 1611-1621 (1997)
MEDLINE 97261963
PUBMED 9108377
AUTHORS

TITLE
Jaguemlin,P., Duviaux,S.M., Jensen,J., Godfraind,C., Gradwohl,G.,
Guillemot,F., Madgen,O.D., Carmeliet,P., Dewerchin,M., Collen,D.,
Roussseau,G.G. and Lemaigre,F.P.
Transcription factor hepatocyte nuclear factor 6 regulates
pancreatic endocrine cell differentiation and controls expression
of the proendocrine gene ngn3
Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
JOURNAL 20285449
MEDLINE 10825208
PUBMED 10825208
REFERENCE
AUTHORS Gradwohl,G.J.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,
CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,
F-67404 ILKIRCH cedex, FRANCE
REMARK Revised by [4]
REFERENCE 4 (bases 1 to 5567)
AUTHORS Lemaigre,F.P.

```

```

TITLE Direct Submission
JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic
Research Unit, Louvain University Medical School, Avenue Hippocrate
75, box 7529, Brussels 1200, BELGIUM
COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
FEATURES
source
1..5567
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="SV129"
/db_xref="taxon:10090"
/clone="lambda GEM12 clone 19c"
/cell_line="D3"
/cell_type="embryonic stem cells"
/clone_lib="genomic library No.62, MboI partial digests
cloned in BamHI"
/dev_stage="embryos"
958
/misc_binding
/bound_molec="HNF6"
3692
/misc_binding
/bound_molec="HNF6"
4145..5567
/gene="MATH4B"
TATA_signal
4145..4151
/gene="MATH4B"
CDS
4923..5567
/note="putative"
/gene="MATH4B"
/codon_start=1
/product="neurogenin 3"
/protein_id="CAA70366.1"
/db_xref="GI:1666088"
/db_xref="SWISS-PROT:P70661"
/translation="MAPK14/POVSPENOPFPASDHVLSNSTPSPPLIPR
DCEAEVDCRGTRKRLARRGRNRRKSLALSKORRRKRRKANDRRPMLNSA
LDALRGVLPTRDDAKLKIETLRFAHYTIALYOTRIADHSFYGPPEPPCELS
PGGSNDGWSIYSPVQAGNLSPASLEPFGVPSPSYLLPGLVPSDFL"

BASE COUNT 1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match 56.2%; Score 819.8; DB 10; Length 5567;
Best Local Similarity 87.1%; Pred. No. 2e-173;
Matches 997; Conservative 0; Mismatches 102; Indels 46; Gaps 7;
Qy 4 GGTAGCGAGAGAGAGAGCCCTGGGCGCCCGCTGCTGATGGCCCGTGCACAGGACGA 63
Db 4424 GGGCTGCGTGCGAGCGCCCTGGGCGCCCGCTGCTGATGGCCCGTGCACAGGACGA 4483
Qy 64 GCCCGGAGGAGCGCTCTGTGTC-CGGGACAGAGAGATTAAGCGTGCAGGAGGACACAG 122
Db 4484 GCCCGGAGGAGCGCTCTGTGTCGGGCGGCGGACAGAGATTAAGCGTGCAGGAGGACACAG 4543
Qy 123 ATT---AGCACTCAGAAATTCCTCTGGGTCTCACACTGCA-CAGAGGCCGAGAGACC 177
Db 4544 ACTTCAGTCAGAGCTCAGAAATTCCTCTGGGTCTCATCTCAGACAGTGTGAGTACT 4603
Qy 178 CCTCGAGCTCTTTGCGCGCCCTCAGAGCAATTAATCTCAGGCGAGGCGCTGCAGCT 237
Db 4604 CTTGCGAGCTTTTTCACACTTTCAGAGCAATTAATCTCAGGCGAGGCGCTGCAGCT 4663
Qy 238 CAGCAAACTTGAAGCAGAGGAGGAGGAGTATCAACCGCTGCTTGACTTGACC- 296
Db 4664 TAGCAAACTTGAAGGAGGAGGAGGAGTATCAACCGCTGCTTGACTTGACC 4723
Qy 297 -----ACCGAGCTCTCTGTTCTTTTGAAGCCCGGAGTA 330
Db 4724 TATCCACTGCTGCTTGTCTACTGACTGACTGCTGCTCTCTTCTTTTGAAGTGGGAG-A 4782
Qy 331 ACTAGTAACATTAGAAGCTCCAAAGGTAAGAAGGAGAGT-----GGTGG 380
Db 4783 ACTAGTAACATTAGAAGCTCCAAAGGTAAGAAGGAGAGTGGGCTGTGTGGG 4842
Qy 381 GCGTACTGATGCCGCGTGAAGTGAAGCTTAAGTCAAGAGACTGTCAACACCC--CCTTC 438

```



```

Db      4843  GGAATCTGTGTCCTCCCGTGAAGTAACTTAAGCAAGGCTGGACACACACACCTTC 4902
Qy      439  CATTTTTCCTCCAACTCAGAGATGGGCTCATCTTGGATGCGGCCAATCCAAAGT 498
Db      4903  CATTTTTCCTCCAACTCAGAGATGGGCTCATCTTGGATGCGGCCAATCCAAAGT 4962
Qy      499  CCAAGAGACCCAGCAACCTTTTCCCGAGCTCCGACCCAGCAAGTGTCTCAATT 558
Db      4963  CCCCAGACACACAACTTTTCCCGAGCTCCGACCCAGCAAGTGTCTCAATT 5022
Qy      559  CCAACCCACTGACCTCCTCTGTAAGGAGGAGTCTCCGAGCAGAGAGAGTACT 618
Db      5023  CCACCCCACTGACCTCCTCTGTAAGGAGGAGTCTCCGAGCAGAGAGGAGTACT 5082
Qy      619  GCCAGGAGACATCCAGAGAGCTCCGAGCTCCGAGGAGGAGGAGGAGGAGGAG 678
Db      5083  GCCAGGAGACCTCCAGAGAGCTCCGAGCTCCGAGGAGGAGGAGGAGGAGGAG 5142
Qy      679  AGTTGGCACTGAGCAGAGAGCAGACCGGCGCAAGAGGCAACGACCGGAGCGCA 738
Db      5143  AGTTGGCACTGAGCAGACAGAGAGCAGACCGGCGCAAGAGGCAATGATCGGAG 5202
Qy      739  ACCGATCACTCACTTAATCTCCGCTGATGCGCTGCGGCTGCTCCCACTTCC 798
Db      5203  ATCCGATCACTCACTTAATCTCCGCTGATGCGCTGCGGCTGCTCCCACTTCC 5262
Qy      799  CGGATGAGGCGCAATTTCAAGAGTGAAGACCTCCGCTGCGCCCAACTAATTGG 858
Db      5263  CGGATGAGGCGCAATTTCAAGAGTGAAGACCTCCGCTGCGCCCAACTAATTGG 5322
Qy      859  CACTGACTCAGACGCTGCGCAGAGGAGCACAAGCTTCTACGCGCCGAGCCCTG 918
Db      5323  CACTGACTCAGACGCTGCGCAGAGGAGCACAAGCTTCTACGCGCCGAGCCCTG 5382
Qy      919  CCTGTGGAGAGTGGGAAAGCCCGGAGGAGGAGGCTCAAGCGGAGCTGAGGAG 978
Db      5383  CCTGTGGAGAGTGGGAGAGCCCGGAGGAGGAGGCTCAAGCGGAGCTGAGGAG 5442
Qy      979  CCCGAGTTTCCCAAGCTGGTAGCTGAGCAGCCCAAGCTCATTTGAGAGAGTTCT 1038
Db      5443  CCCGAGTTTCCCAAGCTGGTAGCTGAGCAGCCCAAGCTCATTTGAGAGAGTTCT 5502
Qy      1039  TGCAGGTCCCAAGCTCCCATCTCTATCTGCTCCGAGGAGCAGCTGATGTTCTCA 1098
Db      5503  TGCAGGTCCCAAGCTCCCATCTCTATCTGCTCCGAGGAGCAGCTGATGTTCTCA 5562
Qy      1099  TGTGA 1103
Db      5563  TGTGA 5567

RESULT 9
AX698801      861 bp      DNA      linear      PAT 02-APR-2003
LOCUS      AX698801
DEFINITION      Sequence 7 from Patent WO02086107.
ACCESSION      AX698801
VERSION      AX698801.1 GI:29499589
KEYWORDS
ORGANISM      Mus musculus (house mouse)
SOURCE      Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 Wobus, A.M., St-Onge, L., Blyszczuk, P. and Hoffmann, U.
            A method for differentiating stem cells into insulin-producing
            cells
JOURNLM      Patent: WO 02086107-A 7 31-OCT-2002;
            Deutscher Akademischer Austauschdienst fuer Entwicklungsbiologische Forschung
            (DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DE)
FEATURES
            source      1..861
            location/Qualifiers

```

```

CDS
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
160..804
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD87442.1"
/db_xref="GI:29499590"
/translation="MAPHPLDALTIQVSPETQPFPGASDHEVLSSNTPSPILIPR
DCEAEVGDORGTGRKLRRGRNRKSRKSLAKSRKSRKSRKSRKSRKSRKSRKSRK
LDARGLVLPFPDPAKLTKEITLFAFNYIALVOTLRADHSRYGEPYPCPELGS
PGGSSNDWMSIYSPVQAGNLSPLAGLEPFGVPSPSYILPLGALVDFL"
BASE COUNT      182 a      274 c      250 g      155 t
ORIGIN
Query Match      46.7%; Score 682.4; DB 6; Length 861;
Best Local Similarity 89.7%; Pred. No. 1.6e-142;
Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;
Qy      312  TTCTTTTGGAGCCCGAGTAATAGTAACATTTAGAACTCCAAAGGATGAGAGGG 371
Db      2    TTCTTTTGGAGCGGAG-ACCTAGTAACATTTGAAACCTCAAAGGATGAGAGGG 60
Qy      372  AGTGGT-----GGGCTACTAGTCCCGCGTGAAGTAACTTAAGTCAGAGA 421
Db      61  CGCGCGGGGTGTGTGTGGGGATTACTGTGTCCTCCCGGAGTAACTTAAGTCAGAG 120
Qy      422  CTG--TCACACCCCTTCCATTTTTCCTCAACCTCAAGATGGGCGCTCATCCCTGGAT 479
Db      121  CTGGACACACACACTTCCATTTTTCCTCAACCGAGATGGGCGCTCATCCCTGGAT 180
Qy      480  GCGCCACCATTCAGAGTGTCCCAAGAGACCCAGCAACCTTTCCCGAGCCTCGAGCAC 539
Db      181  GCGCTCACCATTCAGAGTGTCCCAAGAGACCAACACCAACCTTTCCCGAGCCTCGAGCAC 240
Qy      540  GAAGTGCTCAATTCAATTCACCCCACTAGCCCACTCTGTGTAACGAGGAGTGTCTC 599
Db      241  GAAGTGCTCAATTCAATTCACCCCACTAGCCCACTCTGTGTAACGAGGAGTGTCTC 300
Qy      600  GAAACGAGAGAGGTGATGCTGCGGAGGAGCATTCGAGAGAGTCTCCGCGCGCGGAGAG 659
Db      301  GAAACGAGAGAGGTGATGCTGCGGAGGAGCATTCGAGAGAGTCTCCGCGCGCGGAGAG 360
Qy      660  CGCAACAGGCGCAGAGAGGATGAGCACTGAGCAAGCAGCAGAGAGCGGCGCAAGAG 719
Db      361  CGCAACAGGCGCAGAGAGGATGAGCACTGAGCAAGCAGCAGAGAGCGGCGCAAGAG 420
Qy      720  GCCAAGCAGCGGAGAGCGCAACCGATGCAACCTTAATCTCCGCGCTGATGCGCTGCGC 779
Db      421  GCCAATGATGGGAGAGCGCAATCGCATGCAACCTTAATCTCCGCGCTGATGCGCTGCGC 480
Qy      780  GGTGTCTGCGCCACCTTCCGAGTGAAGCGCAAACTTCAAGAGATCGAGACCTCGCGCTTC 839
Db      481  GGTGTCTGCGCCACCTTCCGAGTGAAGCGCAAACTTCAAGAGATCGAGACCTCGCGCTTC 540
Qy      840  GCCCAACATCAATTTGGGAGCTGACTCAGAGCGTGGCGATAGAGGAGCAACAGCTTAC 899
Db      541  GCCCAACATCAATTTGGGAGCTGACTCAGAGCGTGGCGATAGAGGAGCAACAGCTTAC 600
Qy      900  GGGCCGAGCCCTCTGTGCTCTGTGGAGAGCTGGAGAAAGCCCGGAGGAGGAGCTCCAGCGCG 959
Db      601  GGGCCGAGGAGCCCTCTGTGCTCTGTGGAGAGCTGGAGAAAGCCCGGAGGAGGAGCTCCAGCGCG 660
Qy      960  GACTGGGCTTATCTACTCCCGAGTTTCCCAAGCTGTAGCTGAGCCCGCAAGCTCTCA 1019
Db      661  GACTGGGCTTATCTACTCCCGAGTTTCCCAAGCTGTAGCTGAGCCCGCAAGCTCTCA 720
Qy      1020  TTGAGAGAGTCCCTGGCTGAGAGTGGCGAGCTCCCATCTGTGCTCCCGGAGCAC 1079
Db      721  TTGAGAGAGTCCCTGGCTGAGAGTGGCGAGCTCCCATCTGTGCTCCCGGAGCAC 780
Qy      1080  CTGTGTCTCAGACTTCTGTGAAGGAGCCCAAGGAGCTGGAGGAGGAGCTGGCA 1139

```

Db 781 CTGGTCTCTAGACTTCTTGTGAAGAGACCTGTGCTGTGGTGTGGTCTTAGTG 840

QY 1140 GAAGGAGGAGGAGCTAGAGC 1159

Db 841 GAAAGGAGGAGGAGCAGAC 860

RESULT 10

LOCUS MMU76208 861 bp DNA linear ROD 05-FEB-1997

DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.

ACCESSION U76208

VERSION U76208.1 GI:1815654

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 861)

AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.

TITLE neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS

JOURNAL Mol. Cell. Neurosci. 8 (4), 221-241 (1996)

MEDLINE 97153565

PUBMED 9000438

REFERENCE 2 (bases 1 to 861)

AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-1996) Biology 216-76, California Institute of Technology, Howard Hughes Medical Institute, Pasadena, CA 91125, USA

COMMENT On Feb 5, 1997 this sequence version replaced gi:166911.

FEATURES

source location/Qualifiers

1..861

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

160..804

/gene="ngn3"

160..804

/gene="ngn3"

/function="BHLH transcription factor"

/note="NGN3"

/codon\_start=1

/product="neurogenin 3"

/protein\_id="AAC53029.1"

/translation="MAPHPLDALTIQVPEPTQPPPGASDHEVLSSNSTPSPFLIPR DCEAEVDDCRGTSRKLRRGRGRNRPKSEILSLSKORSRRKADRRNRHNNLSA LDAQGVLPFPDPAKLTKEIETLRANHYIALITOTLRADHSFPGPRPVPCCGLGS PGGSGNDGSLSPVSQGNLSPTASLEPFLDVPSPSYLPLGALVPSDFL"

CDS

gene

BASE COUNT 182 a 274 c 250 g 155 t

ORIGIN

Query Match 46.7%; Score 682.4; DB 10; Length 861;

Best Local Similarity 89.7%; Pred. No. 1.6e-142;

Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;

QY 312 TTTCTTTGAGCGCGGAGTAAGTAAGTATTAAGAACTCCAAAGGTAAAGGGG 371

Db 2 TTTCTTTGAGCGCGAGG-AACTAGTAAGTAAGTATTAAGAACTCCAAAGGTAAAGGGG 60

QY 372 AGTGGGT-----GGGCGTACTCTAGTCTCCGCGTGAAGTGAAGTCTTAAGTCAGAGA 421

Db 61 CGCGCGGGGTGTGTGTGGGAGTACTCTGTCTCCCGGTGAGTGAAGTCTTAAGTCAGAGG 120

QY 422 CTG--TCACACCCCTCTTCAATTTTCCCAACTCAGATGCGCGCTCATCCCTTGAT 479

Db 121 CTGGACACACACACCTTCAATTTTCCCAACCGAGATGCGCGCTCATCCCTTGAT 180

QY 480 GGGCCACCATCAAGTGTCCCAAGAGACCCCTTCCCGGAGCTCGACAC 539

Db 181 GCGCTACCATTCAGAGTCTCCCAAGAGACACACAAACCTTTTCCGAGCCTCGACAC 240

QY 540 GAAGTCTCAGTTCATTCATTCACCCACCTGACCCCATCTGTACGAGGAGCTGTCC 599

Db 241 GAAGTCTCAGTTCATTCATTCACCCACCTGACCCCATCTGTACGAGGAGCTGTCC 300

QY 600 GAAGCAGAGAGGTGACTGCGCAGAGGACATCGAAGAGTCTCGTCCGCGCGAGAGG 659

Db 301 GAAGCAGAGAGGTGACTGCGCAGAGGACATCGAAGAGTCTCGTCCGCGCGAGAGG 360

QY 660 CGCAACAGGCGCCAGAGGAGTGGCACTGAGCAAGCAGCAGACGCGCGCAGAGAG 719

Db 361 CGCAACAGGCGCCAGAGGAGTGGCACTGAGCAAGCAGCAGACGCGCGCAGAGAG 420

QY 720 GCCAACGACCGGAGCGCAACCGCATGACCAACTTAACTCGCGCTGATGCGTGGC 779

Db 421 GCCATGATCGGAGCGCAATCGATGACCAACTTAACTCGCGCTGATGCGTGGC 480

QY 780 GGTGTCTGCGCCACCTTCCCGATGACCGCAAACTTAACTCGCGCTGATGCGTGGC 839

Db 481 GGTGTCTGCGCCACCTTCCCGATGACCGCAAACTTAACTCGCGCTGATGCGTGGC 540

QY 840 GCCCAACTCACTTTGGGCACTGACTGACGCTGCGCATAGCGGACCAAGCTTTAC 899

Db 541 GCCCAACTCACTTTGGGCACTGACTGACGCTGCGCATAGCGGACCAAGCTTTAC 600

QY 900 GGGCCGAGCGCCCTGTGCTGCGGAGGAGTGGGAGCGCGGAGGCGGCTTCCAGCGG 959

Db 601 GGGCCGAGCGCCCTGTGCTGCGGAGGAGTGGGAGCGCGGAGGCGGCTTCCAGCGG 660

QY 960 GACTGGGCTCTATCTATCTCCCAAGTTTCCCAAGCTGTAGCTGAGCCCAAGCTTCA 1019

Db 661 GACTGGGCTCTATCTATCTCCCAAGTTTCCCAAGCTGTAGCTGAGCCCAAGCTTCA 720

QY 1020 TTGAGAGATTCTCTGCTGAGAGGCGCCAAACAGGCGCTGAGGCGCTGAGCA 1079

Db 721 TTGAGAGATTCTCTGCTGAGAGGCGCCAAACAGGCGCTGAGGCGCTGAGCA 780

QY 1080 CTGGTCTCTAGACTTCTTGTGAAGGCGCCAAACAGGCGCTGAGGCGCTGAGCA 1139

Db 781 CTGGTCTCTAGACTTCTTGTGAAGGCGCCAAACAGGCGCTGAGGCGCTGAGCA 840

QY 1140 GAAGGAGGAGGAGCTAGAGC 1159

Db 841 GAAAGGAGGAGGAGCAGAC 860

RESULT 11

AF234829 5340 bp DNA linear PRI 19-OCT-2001

LOCUS Homo sapiens neurogenin 3 gene, complete cds.

DEFINITION AF234829

ACCESSION AF234829.1 GI:13183002

VERSION

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

REFERENCE 1 (bases 1 to 5340)

AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus. Iwasaki, N., Ogata, M., Iwamoto, Y., Schwarz, P.E., Cox, N.J., Mutations in the coding region of the neurogenin 3 gene (NEUROG3) are not a common cause of maturity-onset diabetes of the young in Japanese subjects

TITLE Diabetes 50 (3), 694-696 (2001)

JOURNAL

MEDLINE 21140923

PUBMED 11246894

REFERENCE 2 (bases 1 to 5340)

AUTHORS Lin, J. and German, M.

TITLE Direct Submission

JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave., San Francisco, CA 94143-0534, USA

FEATURES  
source  
Location/Qualifiers  
1. 5340  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
3022..3666  
/product="neurogenin 3"  
/codon\_start=1  
/product="neurogenin 3"  
/protein\_id="AAK15022.1"  
/db\_xref="GI:13183003"  
/translation="MTPQSGAPVQVTRERSEFPRASEDEVCTPSAPSPRTNRG  
NCABEBCGCRGAPKLRARGRSRKPSKSELALSKORSRKANDERNMNLNSA  
LDALRGVLPTEPDADKLKLETRAPAHYIWALETOTRIADHSYLALEPPAPHCSEIG  
SPGSGPDMGSLVSPVSGSLSPASLEERPLLIGATISACLSPGSLAFSPFL"

CDS  
3022..3666  
/product="neurogenin 3"  
/codon\_start=1  
/product="neurogenin 3"  
/protein\_id="AAK15022.1"  
/db\_xref="GI:13183003"  
/translation="MTPQSGAPVQVTRERSEFPRASEDEVCTPSAPSPRTNRG  
NCABEBCGCRGAPKLRARGRSRKPSKSELALSKORSRKANDERNMNLNSA  
LDALRGVLPTEPDADKLKLETRAPAHYIWALETOTRIADHSYLALEPPAPHCSEIG  
SPGSGPDMGSLVSPVSGSLSPASLEERPLLIGATISACLSPGSLAFSPFL"

BASE COUNT 1215 a 1500 c 1514 g 1111 t

ORIGIN

Query Match 35.4%; Score 517.4; DB 9; Length 5340;  
Best Local Similarity 67.4%; Pred. No. 1.3e-105;  
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

20 GTCCCTGGCCCCCGTGTCTGATTGGCCGTTGACAGACAGACAGCCGCGACGCGCT 79  
2535 GGCCCGACAGGCGCCGCGCTGATTGGCGGTGGCGCGGACAGCCGGGACGACGCT 2534  
80 CCTGTCCGGGACAGACAGATTAAGCTGTCACAGGAGACACAGATTAGCAGTGAAGT 139  
2595 CCTGGCCGGGCGAAGAGATTAAGCTGTCACAGGAGACACAGATTAGCAGTGAAGT 2654  
140 CCTGTGGGCTCACACATGC-----ACAGAGCGGAGGACCCCTC 181  
2655 TCCCTGGGCTCACACGCGCGCTGACAGAGACGTCAGAGAGCCCTCGGACCCCTT 2714  
182 CGAGCTTCTTGTGCTCCAGACGCAATTACT-CCAGGCGAGGCGCTGACAGCTCAG 240  
2715 CTCTCTTCTTCTTCTTGTGGGGCTGGGCAACTCCAGCGGGGCGCTGACAGCTCAG 2774  
241 CAAACTTCGAAGCAGCAG-----AGGGGTTCAGTATCCACCGCTGCT 285  
2775 CTGAATCTGGGACCGAAGCGCGCTGAGCTCCACGCGCTCTCTCTAGCTGCTC 2834  
286 TGACTCTGACACCGGACGCTCTCTGTTCTTTTGAAGCGGAGTAACTAGGTAACATTTA 345  
2835 TATCTTTTGGCGCGGTGAAGAGTAATTTTGAAGGCTCTCCAGAGGACGGGACGGGAA 2894  
346 GGAAC-CTCCAAAGGGTAGAAGAGGGAGTGGGTGGGCTACTAGTCCGCGTGAGT 404  
2895 AGAGGATCTCTGTGACCGAGCGGGGCTGGAGATGCTGTTTTTTTCCCACT 2954  
405 GACTCTTAAGTCAGAGACTG-----TACACCCCCCTTCCATTTTTCCTCA 451  
2955 AGCTTCGAATCGCGAGCTGGCGCTGACGGAAGCTCAACCTTCCCTCTGACCCCG 3014  
452 CCTAGAGATGGGCTCATCCCTTGAATGGCGCCACATCCAGAGTCCCAAGAGACCA 511  
3015 CCGTAGATGAGGCTCAACCTCGGGTGGCGCCACTGTCCAGAGTACCGGTGAGACGA 3074  
512 GCAACCTTTTCCGAGAGCTGAGACCAAGAGTCTCAATTCCATTCACCCCACTAG 571  
3075 GCGGCTTTCCTCCAGAGCTCGGAAGAGAGTAGTACCTGCCACGTCCGCCCGCCAG 3134  
572 CCCCACTCTGTAACCGAGGAGTCTCCGAAGACAGAGAGTGACTGCGGAGGACATC 631  
3135 CCCCACTGCAACCGGAGGAACTGCGGAGAGCGGAGAGAGGAGGCTCCAGAGGAGCCC 3194  
632 GAGGAGTCTGCGGCGGAGGCGGAGGCGGACAGGCGCCAAAGCGAGTGGACATGAG 691  
3195 GAGGAGTCTCGGAGGAGGCGGAGGCGGAGGCGGCGCTTAAGACGAGTGGACATGAG 3254  
692 CAAGCAGCAGAAACCGGCGCAAGAGGCGCAACGACCGGAGCGCAACCGCATGCAAA 751

Db 3255 CAAGCAGCAGAGGATGCGCGAAGAAAGGCCCAACGACCGGAGCGCAATGATGCAAA 3314  
Oy CCTAACTCCGCGGTGATGGCTGCGCGGATGCTCTGACCACTTCCCGATGACGCCAA 811  
752 CCTAACTCCGCGGTGATGGCTGCGCGGATGCTCTGACCACTTCCCGATGACGCCAA 811  
Db CCTAACTCCGCGGTGATGGCTGCGCGGATGCTCTGACCACTTCCCGATGACGCCAA 3374  
Oy 812 ACTTACAAAGATGAGACCTTGGCTTCCGACCACTTACATTTTGGCATGACTGACAG 871  
Db 3375 GCTCACAAGATGAGACCTTGGCTTCCGACCACTTACATTTTGGCATGACTGACAG 871  
Oy 872 GCTCACAAGATGAGACCTTGGCTTCCGACCACTTACATTTTGGCATGACTGACAG 871  
Db 3435 GCTCACAAGATGAGACCTTGGCTTCCGACCACTTACATTTTGGCATGACTGACAG 871  
Oy 932 GGAAGCGCGGAGGAGGCTCCAGCGGAGCTGAGGAGCTTATCTACTCCCACTTCCCA 991  
Db 3495 GCTGAGGAGCGGAGGAGGCTCCAGCGGAGCTGAGGAGCTTATCTACTCCCACTTCCCA 991  
Oy 992 AGCTGTAGCCTGAGACCCCAAGCCTCATTTGAGAGAGTCCCTGAGTGGCCAG 1051  
Db 3555 GCTGAGGAGCGGAGGAGGCTCCAGCGGAGCTGAGGAGCTTATCTACTCCCACTTCCCA 991  
Oy 1052 CTCCCACTCTGCTGCTCCCGGAGACCTTGTGTTCTGACACTTCTTGAAGGAGCCCA 1111  
Db 3615 CTCTTCCGCTGCTTGAAGCGGAGCACTTGTGTTCTGACACTTCTTGAAGGAGCT 3674  
Oy 1112 AAGAGCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1171  
Db 3675 GTCTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3734  
Oy 1172 AAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196  
Db 3735 AGGTTGCGCAGCGGCGGCGGCGCTC 3759

RESULT 12  
AL450311/C  
LOCUS  
DEFINITION  
Human DNA sequence from clone Rp11-3433 on chromosome 10, complete sequence.  
ACCESSION  
AL450311  
VERSION  
AL450311.11 GI:14626972  
KEYWORDS  
HTG.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
REFERENCE  
Bukariyola; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 165110)  
AUTHORS  
Howden, P.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT  
On Jul 8, 2001 this sequence version replaced gi:14575291.  
During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPeP; Information on the WormPeP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human

Chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>  
Rp11-343J3 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6  
This sequence is the entire insert of clone Rp11-343J3 The true left end of clone Rp11-242G20 is at 139955 in this sequence. The true right end of clone Rp11-404C6 is at 6588 in this sequence.

## FEATURES

source  
1. .165110  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="Rp11-343J3"  
/clone\_11b="RPCT-11.2"  
7. .147  
/note="THE1C repeat: matches 2. .142 of consensus"  
repeat\_region  
151. .293  
/note="MIR repeat: matches 2. .148 of consensus"  
repeat\_region  
1136. .1187  
/note="26 copies 2 mer gt 98% conserved"  
repeat\_region  
2960. .3272  
/note="AluSg repeat: matches 1. .310 of consensus"  
repeat\_region  
4144. .4270  
/note="AluX repeat: matches 1. .114 of consensus"  
repeat\_region  
4319. .4494  
/note="AluX repeat: matches 118. .293 of consensus"  
repeat\_region  
5662. .6438  
/note="L1MEC repeat: matches 272. .1095 of consensus"  
repeat\_region  
6916. .8080  
/note="L1MEC repeat: matches 1168. .2267 of consensus"  
repeat\_region  
9934. .10087  
/note="MIR repeat: matches 25. .184 of consensus"  
repeat\_region  
10353. .10481  
/note="MIR repeat: matches 65. .194 of consensus"  
repeat\_region  
11077. .11372  
/note="AluX repeat: matches 1. .294 of consensus"  
repeat\_region  
11600. .11639  
/note="20 copies 2 mer tc 95% conserved"  
repeat\_region  
12479. .12608  
/note="65 copies 2 mer at 90% conserved"  
repeat\_region  
13433. .13876  
/note="MIR repeat: matches 1. .466 of consensus"  
repeat\_region  
14372. .14576  
/note="AluSg repeat: matches 1. .301 of consensus"  
repeat\_region  
14577. .14633  
/note="AluSg repeat: matches 243. .299 of consensus"  
repeat\_region  
14843. .14943  
/note="MIR repeat: matches 9. .109 of consensus"  
repeat\_region  
15404. .15613  
/note="MIR repeat: matches 3. .213 of consensus"  
repeat\_region  
16626. .16705  
/note="MIR repeat: matches 60. .140 of consensus"  
repeat\_region  
16928. .16993  
/note="33 copies 2 mer gg 66% conserved"  
repeat\_region  
17476. .17569  
/note="MIR repeat: matches 2. .114 of consensus"  
repeat\_region  
17719. .18069  
/note="L1MC4 repeat: matches 7617. .7977 of consensus"  
repeat\_region  
18117. .18345  
/note="MIR repeat: matches 7. .262 of consensus"  
repeat\_region  
19169. .19400  
/note="MIR repeat: matches 5. .239 of consensus"  
repeat\_region  
19436. .19627  
/note="L1MC5 repeat: matches 7728. .7917 of consensus"  
repeat\_region  
19642. .19716  
/note="MIR repeat: matches 48. .131 of consensus"  
repeat\_region  
19969. .20010  
/note="21 copies 2 mer tg 100% conserved"  
repeat\_region  
20623. .20704

/note="41 copies 2 mer gt 85% conserved"  
repeat\_region  
20738. .20773  
/note="U2 repeat: matches 1. .36 of consensus"  
repeat\_region  
20802. .20863  
/note="MIR repeat: matches 1. .62 of consensus"  
repeat\_region  
20889. .21263  
/note="MIR repeat: matches 188. .541 of consensus"  
repeat\_region  
21463. .21618  
/note="MIR repeat: matches 46. .192 of consensus"  
repeat\_region  
22019. .22326  
/note="AluY repeat: matches 1. .306 of consensus"  
repeat\_region  
22381. .22564  
/note="U2 repeat: matches 2453. .2629 of consensus"  
repeat\_region  
22896. .23174  
/note="MIR repeat: matches 117. .413 of consensus"  
repeat\_region  
23215. .23346  
/note="MIR repeat: matches 48. .188 of consensus"  
repeat\_region  
23388. .23532  
/note="U2 repeat: matches 2097. .2230 of consensus"  
repeat\_region  
23533. .23837  
/note="AluY repeat: matches 1. .303 of consensus"  
repeat\_region  
23838. .24137  
/note="U2 repeat: matches 1754. .2097 of consensus"  
repeat\_region  
24251. .24581  
/note="AluX repeat: matches 1. .300 of consensus"  
repeat\_region  
24653. .24850  
/note="MIR repeat: matches 1. .200 of consensus"  
repeat\_region  
26189. .26685  
/note="CpG island"  
/evidence="not\_experimental"  
29453. .30918  
/note="CpG island"  
/evidence="not\_experimental"  
31084. .31156  
/note="MIR repeat: matches 65. .138 of consensus"  
repeat\_region  
31618. .31876  
/note="AluY repeat: matches 29. .275 of consensus"  
repeat\_region  
32767. .32830  
/note="MIR repeat: matches 76. .139 of consensus"  
repeat\_region  
33050. .33178  
/note="43 copies 3 mer tcc 72% conserved"  
repeat\_region  
35112. .36201  
/note="CpG island"  
/evidence="not\_experimental"  
36551. .36604  
/note="27 copies 2 mer ac 94% conserved"  
repeat\_region  
37585. .38254  
/note="L1PA15 repeat: matches 5480. .6157 of consensus"  
repeat\_region  
38265. .38561  
/note="AluX repeat: matches 1. .300 of consensus"  
repeat\_region  
39214. .39423  
/note="MIR repeat: matches 3. .219 of consensus"  
repeat\_region  
39489. .39801  
/note="AluX repeat: matches 1. .292 of consensus"  
repeat\_region  
40050. .40189  
/note="MIR repeat: matches 1. .144 of consensus"  
repeat\_region  
41057. .41290  
/note="MIR repeat: matches 8. .240 of consensus"  
repeat\_region  
41296. .41460  
/note="U2 repeat: matches 1916. .2416 of consensus"  
repeat\_region  
41506. .41944  
/note="U2 repeat: matches 1448. .1779 of consensus"  
repeat\_region  
42388. .42698  
/note="U2 repeat: matches 1448. .1779 of consensus"  
repeat\_region  
44193. .44579  
/note="THE1C repeat: matches 1. .371 of consensus"  
repeat\_region  
44600. .44956  
/note="L1R16A repeat: matches 90. .445 of consensus"  
repeat\_region  
45240. .45300  
/note="MIR repeat: matches 37. .97 of consensus"  
repeat\_region  
45798. .45909  
/note="L1R11 repeat: matches 90. .192 of consensus"  
repeat\_region  
46826. .46871





	misc_feature	74797..81326	/note="assembly_fragment"
	misc_feature	81427..89446	/note="assembly_fragment"
	misc_feature	89547..96459	/note="assembly_fragment"
	misc_feature	96560..106498	/note="assembly_fragment"
	misc_feature	106599..116986	/note="assembly_fragment"
	misc_feature	117087..128890	/note="assembly_fragment"
	misc_feature	clone_end..T7	
	misc_feature	vector_side..right"	
	misc_feature	128991..147290	/note="assembly_fragment"
	misc_feature	147391..173341	/note="assembly_fragment"
Query Match	35.1%;	Score 512.6;	DB 2; Length 173341;
Best Local Similarity	66.7%;	Prod. No. 1.1e-104;	
Matches	818;	Conservative	0; Mismatches 159; Indels 50; Gaps 4;
QY	20	GTCCCTGGGCCCCCGTTGCTGATTTGGCCCGTGGCACAGCAGCCCGGACGACCGCT	79
Db	7254	GGCCCCAGGCCCCCGGCGCTGATTTGGCCCGTGGCACAGCAGCCCGGACGACCGCT	7199
QY	80	CCTGGTCGGGCGAGCAGATTAAGCCTGACGAGGAGCACACGATTAGACTCAGAAAGT	139
Db	7194	CTGCGCCGGGCGAAGCAGATTAAGCGTGCAGAGGGGACACAGACTTGCTCAGAAA	7135
QY	140	CCCTTGGGCTTCACCACTGC-----ACAGAGCCGAGAGACCCCTC	181
Db	7134	TCCCTGCGGTCTCACCGCCGCGCCTCGAGAGAGCCTGACACAGAGCCTCGGACCCCA	7075
QY	182	CGAGCTTTTGCTGCTCCCTCCAGACGCAATTTACT--CGAGGGGAGGGGCGCTGACACTAG	240
Db	7074	CTCTCTCTTTTCTCTTTTGGGGCTGGGGCAATCCAGGGCGGGGCGCTGCACTGAG	7015
QY	241	CAAACTTCGAAGCAGCAG-----AGGGGTTCACTATTCACCGCTGCT	285
Db	7014	CTGAACCTTGGGAGCAGCAAGAGCCCGCTGAGCTCCCAACGCGCTGCTCATGCTCTC	6955
QY	286	TGACTCTGACCAACCGCAGCTCTCTGTTCTTTTGACCCTCGAGATTAAGTAACTATTA	345
Db	6954	TATCTTTTGGCGCGGTGAAAGATTAATTTTGACAGCCTCCGAGGAGCGGGACGGGAA	6895
QY	346	GGAACCTCCAAAGGATAGAAAGGGGAGTGGGTGGCGTACT-----CT	389
Db	6894	AGAGGGATCTCTGACCCAGCGGGGGCTGGGAGATGCTGTTTTGTTTTTCCACCT	6835
QY	390	AGTCCCGGTGGAAGTACCTCAAGTCAAGACTGACACACCCCTTCATTTTTC	449
Db	6834	AGCTTCGGAATCGCGGACTGCGGCCCAATGACGGAATCAAACTTACCTTCCCTCTGACCC	6777
QY	450	AACCTCAGATGCGCGCTCATCTCTTGATGCGCCACACATCAAGTATCCCAAGAGAC	509
Db	6774	CGCGTAGAATGAGCGCTCAACCTCGGGTGGCCCACTGTCATAGTGACCGGTAGACG	6718
QY	510	CAGCAACCTTTCCCGGAGCTTCGACCAACGAAATGCTCAGTTCCAAATTTCAACCCCACT	569
Db	6714	GAGCGGTCTTTCCCAAGAGCTTCGAAAGCAGAAATGACCTGCCCCACGTCGCCGCCGCC	6655
QY	570	AGCCCACTTCGTACCGAGGGAGTGTCCGAAGAGAGAGCAGTGACTGCGCAGGAGCA	629
Db	6654	AGCCCACTTCGACACCGGGGAATCTGCGAGCGGAAAGAGAGGCTGTCGAGAGGCGC	6599
QY	630	TCGAGGAAGTCTCGTGCAGCGCGCGCGAGAGGCGCAACAGGCCAAGAGCGAGTTGGCACTG	689
Db	6594	CGAGGAAGTCTCGGGGCAACGGCGGGGGGACCAACCGGCTTAAAGGAGATTTGGCACTG	6533
QY	690	AGCAAGACGACGAAGCGCGGTGCAAGAAAGTCAACGACCGGGAGCGCACTGCATGCAC	749

Db	Accession	Gene	Chromosome	Start	End	Strand	Orientation	Feature
Db	6534	AGCAAGCAGCAGCGAGTCCGGCGAAAGAGGCGCAACAGCCGCGGAGCGCCAAATCGAATGCAC	10q21.3	1157	1321	+	coding	6475
Qy	750	AACTTAACTCCGGCTGATGCGCTCCGGGATGTTCTTGGCCACACTTCCCGGAGAGACGCC	10q21.3	1157	1321	+	coding	809
Db	6474	AACTTCAACTCGGACCTGAGACGCCCTTCGGGATGTTCTTGGCCACACTTCCCGGAGAGACGCC	10q21.3	1157	1321	+	coding	6415
Qy	810	AAACTTCAAAAGATCGAGACCCCTGCGCTCGCCCAACTATCATTTGGGACTGACTCAG	10q21.3	1157	1321	+	coding	869
Db	6414	AAGCTCACACAGATTCGAGACCGTTCGCTTGGCCCAACTTACATCTGGGCGCTACTCA	10q21.3	1157	1321	+	coding	6355
Qy	870	ACGCTGCGCATATGCGGACCAACAGCTTCTTACGAGCCCGAGCCCGCTGTGACCTGTGGGAG	10q21.3	1157	1321	+	coding	929
Db	6354	ACGCTGCGCATATGCGGACCAACAGCTTCTTACGAGCCCGAGCCCGCTGTGACCTGTGGGAG	10q21.3	1157	1321	+	coding	6295
Qy	930	CTGGGAAGCCCGGAGGGGGCTCCAGGCGGACATCTGGGGCTTATCTACTCCCACTTCC	10q21.3	1157	1321	+	coding	989
Db	6294	GAGCTGGGCGAGCCCGAGCGGTTCCCGCGGGGACATGGGGAGTCCCTTACTCCCCAGTCTCC	10q21.3	1157	1321	+	coding	6235
Qy	990	CAAGCTGGTAGCTCGAGGCCCAACAGCCCTCATTTGAGAGAGTTCCCTGCGCTTCAGAGTCC	10q21.3	1157	1321	+	coding	1049
Db	6234	CAGGCTGGACACTGATGATCTCCCGCGCTCGCTGAGAGAGGACACCCGGGCTGCTGGGGCC	10q21.3	1157	1321	+	coding	6175
Qy	1050	AGCTCCCATCTCTGCTGCTCCCGGAGCACCTGCTGTTCTTCAAGCTTCTTGTGAAGGGCC	10q21.3	1157	1321	+	coding	1109
Db	6174	ACCTTTTCCGCTCTTGAAGCCCAAGGACATGCTTGTCTTCTTCAATTTTCTGTGAAGAGAC	10q21.3	1157	1321	+	coding	6115
Qy	1110	CAAAAGAGCCCTGGGGCGGTGGGCGCTGGCAGAAAGGAGGAGTCAAGCTGTCTGAAT	10q21.3	1157	1321	+	coding	1169
Db	6114	CTGTCTGTGCTGGGAGCTGTGGGTCTTAAGGATGAGGAGAGGAGGAGCGGAGCGGT	10q21.3	1157	1321	+	coding	6055
Qy	1170	GGAAGTATGTGAGGCACTGAGATC	10q21.3	1157	1321	+	coding	1196
Db	6054	AGAGGATGGCCGAGCGCGCGCCCTC	10q21.3	1157	1321	+	coding	6028









**THIS PAGE BLANK (USPTO)**





```
FT /product= "Ngn3"
FT /note= "Neurogenin 3"
PN MO200059936-A1.
XX 12-OCT-2000.
XX 28-MAR-2000; 2000WO-US08436.
XX 06-APR-1999; 99US-0128180.
XX (RESC ) UNIV CALIFORNIA.
XX German MS, Lin J;
XX MPI; 2000-664989/64.
XX P-PSDB; AAY85618.
XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX them, useful for diagnosis, prevention and treatment of diabetes
XX mellitus and to identify individuals at risk of diabetes -
XX Claim 18; Page 49-50; 54pp; English.
XX
XX The human neurogenin 3 Ngn3 DNA sequence AAc61089 encodes the Ngn3
XX protein AAY85617. The Ngn3 gene is located at chromosome position
XX 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX protein sequences, and includes an antibody recognising the Ngn3 protein.
XX Also included in the invention is a method for identifying an islet cell
XX precursor, the method involves analysing a cell for the expression of the
XX Ngn3 gene product, where detection of the product is indicative of an
XX islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX reagent for detecting (in a subject) a predisposition to a defect in
XX pancreatic islet cell function or formation associated with a defect in
XX Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX precursor cells expressing Ngn3, and to alter cellular differentiation in
XX culture in vivo to produce new beta-cells to treat patients with diabetes
XX mellitus. The present sequence represents the murine Ngn3 genomic DNA
XX sequence.
XX
XX Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;
XX
XX Query Match 61.7%; Score 900.6; DB 21; Length 1861;
XX Best Local Similarity 86.2%; Pred. No. 4.9e-235;
XX Matches 1094; Conservative 0; Mismatches 129; Indels 46; Gaps 7;
XX
OY 4 GGTACGAGAGAGAGAGTCCCTGGGCCCCCGTGTGATTGGCCCGTGACACAGGACCA 63
DB 594 GGGCTCGCGTGGCGCAGCCCTCGGCCCCCTCCGCTGATTGGCCCGTGTGACAGCA 653
OY 64 GCCCGGAGGACCGCTCTGTGTC -GGGACAGAGCAGATTAAGCCTGCCAGGAGACACG 122
DB 654 GCCCGGAGGACCGCTCTGTGTCGGGGGAGAGAGAGATTAAGCCTGCCAGGAGACACG 713
OY 123 ATT-----AGAGCTCAGAAAGTCCCTCTGGAGTCTACCACTGCA -CAGAGGCCAGAGACC 177
DB 714 ACTTGACATGACACTGAGAAATCCCTCTGGAGTCTATCACTGACAGAGAGTGTGAGTACTT 773
OY 178 CCTCGAGCTTCTTGTGCTGCTCCAGACGCAATTACTCAGGCGAGGGCGCTGACACT 237
DB 774 CCTCGAGCTTCTTGTGCTGCTCCAGACGCAATTACTCAGGCGAGGGCGCTGCAATT 833
OY 238 CAGCAAACTTTCGAAAGCAGAGAGAGGGTTCAAGCTATCCACCGCTCTTGACTTGACC- 296
DB 834 TAGCAGAACTTCAGAGGAGAGAGAGGCTCAGCTATCCACGCTGCTTGAACACGACACC 893
OY 297 -----ACCGCAGCTCTGTGTTCTTTTGAAGCCGAGTA 330
DB 894 TATCACTGCTGCTGTCTGCTACATGACCTGCTGCTCTTATTTCTTTTGAAGCTGAGAG-A 952
OY 331 ACTAGTAACATTTAGAACTTCAAAAGGTAAGAAAGGAGAGTGGT-----GG 380
DB 953 ACTAGTAACATTTAGAACTTCAAAAGGTAAGAAAGGAGAGGAGGAGGCTGTGTGGG 1012
```

```
OY 381 GCGTACTTATGTCCTCCGGTGGAGTACCTCTAAGTCAGAGACTG--TCACACCCCTTTC 438
DB 1013 GGATACCTGAGTCCCGGTGAGTGAACCTTAAGTCAGAGGCTGACACACACACCTTC 1072
OY 439 CATTTTTCCTCACTTCAAGATGAGGCTCTATCCCTTGGATGAGCCCACTCAAGTGT 498
DB 1073 CATTTTTCCTCACTTCAAGATGAGGCTCTATCCCTTGGATGAGCCCACTCAAGTGT 1132
OY 499 CCCAAGAGACCCAGCAACCTTTCGGAGGCTGAGACCAAGAGTGTCAATT 558
DB 1133 CCCAAGAGACCAACACTTTTTCGGAGCTTGAACCAAGAGTGTCAATT 1192
OY 559 CCACCCCACTAGCCCACTCTCGTACCGAGGAGTGTCTCGAAGCAAGACAGTGAAT 618
DB 1193 CCACCCCACTAGCCCACTCTCATCTAGGAGCTGTCTCGAAGCAAGAGTGTGACT 1252
OY 619 GCCGAGGACATCGAGAGAGTCTCGTGGCGGGCGCGGAGGGCGCAACAGCCCAAGAGCG 678
DB 1253 GCCGAGGAGCTCGAGGAGAGTCTCGGCGCCGACCGGAGGGCGCAACAGGCGCAAGAGCG 1312
OY 679 AGTTGGACATGAGAGACGAGAGAGCGGCGGCAAGAGGCCAAGACCGGAGAGCGCA 738
DB 1313 AGTTGGACATGAGAGACGAGAGAGCGGCGGCAAGAGGCCAAGATGAGGAGAGCGCA 1372
OY 739 ACCGATGACCAACCTTAATCCGCGCTGATGAGTGGCGGGTGTCTGCGCCACTTTC 798
DB 1373 ATGCAATGACACACTTCACTCGGCGCTGATGAGTGGCGGGTGTCTGCGCCACTTTC 1432
OY 799 CGAGTGAAGCCCAACTTACAAAGATGAGAGCCCTGCGCTTGGCCACCACTCAATTGGG 858
DB 1433 CGAGTGAAGCCCAACTTACAAAGATGAGAGCCCTGCGCTTGGCCCACTCAATTGGG 1492
OY 859 CACTGACTCAGACCTGCGATGACGAGACCACTTCTAGGCGCCGAGGCCCTGTGTC 918
DB 1493 CACTGACTCAGACCTGCGATGACGAGACCACTTCTAGGCGCCGAGGCCCTGTGTC 1552
OY 919 CCGTGGGAGAGCTGGAAGAGCCGGGAGGGGCTCCAGCGGCGATGAGGGCTCTAATCTACT 978
DB 1553 CCGTGGAGAGCTGGAAGAGCCGGGAGGGGCTCCAGCGGAGCTCTAATCTACT 1612
OY 979 CCCAGATTTCCCAAGCTGTAGCTTGAAGCCCAAGCCCACTTATGAGAGAGTTCCTGCGCC 1038
DB 1613 CCCAGATTTCCCAAGCTGTAGCTTGAAGCCCAAGCCCACTTATGAGAGAGTTCCTGCGCC 1672
OY 1039 TGCAGGTGCCAGTCTCCCATCTGTCTGTCTCCGGGACACCTTGATGTTCTCAACTTCT 1098
DB 1673 TGCAGGTGCCAGTCTCCCATCTGTCTGTCTCCGGGACACCTTGATGTTCTCAACTTCT 1732
OY 1099 TGTGAAGGAGCCCAAGAGAGCCCTGAGGAGGAGCTGAGCAAGAAAGGAGAGTCAAG 1158
DB 1733 TGTGAAGAGACTTGTCTGTGCTGTGGTGTGGTGTGCTTATGTAAGAGAGAGAGCCAGAG 1792
OY 1159 CTGTCTGAATGAGAGGATGAGAGCACTGAGCACTGCGCCCTTCTGAGCTTTCAATTA 1218
DB 1793 CCGTCTGAGAGGAGAGATGAGAGGCTTCAACATCTGCGCTTCTGAGCTTTCACTA 1852
OY 1219 GTCAAGTTC 1227
DB 1853 CTTGAGATCC 1861
RESULT 3
AAD46872
ID AAD46872 standard; DNA; 1860 BP.
XX AAD46872;
XX AC
XX 27-JAN-2003 (first entry)
XX Murine neurogenin 3 (Ngn3) gene.
XX DE
XX Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
```

Query Match	Best Local Similarity	Score	DB	Length
Matches 1093; Conservative	0; Mismatches 129; Indels 46; Gaps 7	61.6%; 86.2%;	899.6; 9.2e-235;	1860;
Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;				
Query	4	GGTAGCGAGAGGAGCGAGTCCCTGCGGCCCCCGTTGCTGATTGGCCCGTGGCACAGGACGA	63.	
Db	594	GGGCTCGCGCTGGCGGCGCCCTGGGGCCCCCTCGCTGATTGGCCCGTGGTGGCGAGGACA	653	
QY	64	GGCCGGCAGGACAGCGCTCTGTGTC-CGGGCGAGACAGATTAAGGGTCCAGAGGACACAG	122	
Db	654	GCCCGGACGACGCTCTGTGCGGGGGGCGAGACAGATTAAGGCTCCAGGGGACACAG	713	
QY	123	ATT-----AGCAGCTCGAAGTCCCTTGGGCTCACCACTGCA-CAGAGCGGAGACCC	177	
Db	714	ACTTGCACTCAGCTCAGAAATCCCTCTGGGTCATCACTGACGACAGTGGTGAATCT	773	
QY	178	CCTCCGAGCTTCTTTGCTGCTCCCTCCAGACCAATTAATCTCCAGGCGAGGGCGCTGACGT	237	
Db	774	CCTCGAGCTTTTCTTCAAGCACTTCCAGACCAATTAATCTCCAGGCGAGGGCGCTGACGT	833	
QY	238	CAGCAAACTTGGAGCGAGACGAGAGGGTTAGCTATCCACCGCTGACTGACCC-296		
Db	834	TGCGAAGACTTCAAGAGGAGACAGAGGCTCACTATCCACTGCTGTTGACCTAACCC	893	
QY	297	-----ACCGCAGCTCTCTGTTCTTTTGGAGCCCGAGATA	330	
Db	894	TATCGACTGCTGCTGTGCTACGTACCTGCTGCTCTCTATTTCTTTTGAAGTGGGAG-A	952	
QY	331	ACTAGGTAACTTTAGAACTTCAAAAGGTTAGAGAGGGAGTGGT-----GG	380	

Db	953	ACTGATGTAACAATTTCGGAACTCCAAAGGGTGGATGAGGGGCGCGCGGGGTGTGTGGG	1012
QY	381	GCGTACTCTAGTCCCCGGTGGAGTGACCTCTAAGTCAGAACTG--TCACACCCCCCTTC	438
Db	1013	GGATACCTGTGATCCCCGGTGCAGTGAACCTCTAAGTCAGAGGCTGGCAACACACACCTTC	1072
QY	439	CATTTCCTCCAACTGAGGATGGGCGCTCAATCCCTTGATATGGGCGCACCATCCAAATGT	498
Db	1073	CATTTTCCTCCAACTGAGGATGGGCGCTCAATCCCTTGATATGGGCGCTCAATCCAAATGT	1132
QY	499	CCCAAGAGACCCAGCAACCTTTTCGCGAGCCTTGAGACAAGAGTCTCAATTC	558
Db	1133	CCCCAGAGACCAACAACTTTTCGCGAGCCTTGAGACAAGAGTCTCAATTC	1192
QY	559	CCACCCCACTTAGCCCCAATCTCTGATCCGAGGAGCTGCTCCGAAGCAAGAGAGTGA	618
Db	1193	CCACCCCACTTAGCCCCAATCTCTGATCCGAGGAGCTGCTCCGAAGCAAGAGAGTGA	1252
QY	619	GCCGAGGGAGCAATGAGGAAGCTCCGTCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGC	678
Db	1253	GCCGAGGGAGCAATGAGGAAGCTCCGTCGCGGCGCGGAGGGCGCAACAGGCCCAAGAGC	1312
QY	679	AGTTGGCACTGAGCAAGCAGCAGCAAGACCGCGCGCAAGAAAGGCCCAAGACCCGGAGCGCA	738
Db	1313	AGTTGGCACTGAGCAAGCAGCAGCAAGACCGCGCGCAAGAAAGGCCCAAGATGATGGGAGCGCA	1372
QY	739	ACGCGATGCAACACTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCTGTCCACTTTC	798
Db	1373	ATCCCAATGCAACACTTAACTCCGCGCTGGATGCGCTGCGCGGTGTCTGTCCACTTTC	1432
QY	799	CGGATGACGCGCAACTTAAGAAAGATGAGAACCTTCGCGCTTCGCCACAACTAATTTGGG	858
Db	1433	CGGATGACGCGCAACTTAAGAAAGATGAGAACCTTCGCGCTTCGCCACAACTAATTTGGG	1492
QY	859	CACATGACTCAGACCGCTGGCATAGCGGACCAAGACTTCTACGAGCCCGAGCCCTGTGC	918
Db	1493	CACATGACTCAGACCGCTGGCATAGCGGACCAAGACTTCTACGAGCCCGAGCCCTGTGC	1552
QY	919	CCTGTGGGAGCTGGGAAGCCCGGGAAGGGGCTCCAGGCGGCACTGGGGCTTAATCTA	978
Db	1553	CCTGTGGGAGCTGGGGAAGCCCGGGAAGGGGCTCCAGGCGGCACTGGGGCTTAATCTA	1612
QY	979	CCCCAGTTTCCCAAGCTGTGTAGCTGAGCCCCCAAGACTCTATTGGAAGAGTTCCTGTGC	1038
Db	1613	CCCCAGTTTCCCAAGCTGTGTAGCTGAGCCCCCAAGACTCTATTGGAAGAGTTCCTGTGC	1672
QY	1039	TGCAGAGTCCAGAGTCCCATCTGATGCTCCCGGGGCACTGTGATCTCAGACTTCT	1098
Db	1673	TGCAGAGTCCAGAGTCCCATCTGATGCTCCCGGGGCACTGTGATCTCAGACTTCT	1732
QY	1099	TGTGAAGGCGCCAAACAGGCCCTTGGGCGGTGGGCGCTGGCAAGAAAGGAGAGTCAAG	1158
Db	1733	TGTGAAGGCGCCCTGTGTGGCTCTGAGGTGTGGGTGATGTAAGAAAGGAGGAGCCAGAG	1792
QY	1159	CTGTCTGAATGGAAGGTATGAGGAGCACTGAGAGATCTCGCCCTTCTGTGCTTTCAATTA	1218
Db	1793	CCGTCTGAGAGGAGGATGATGAGGCTCTCAAGCATCTGCGCTTCTGTGCTTTCACTA	1852
QY	1219	GTCAGGTC 1226	
Db	1853	CTTGGATC 1860	
RESULT 4			
AAAF27254			
ID	AAAF27254 standard; cDNA, 5567 BP.		
XX	AAAF27254;		
AC			
XX	24-APR-2001 (first entry)		
DT			
XX	Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.		

XX	Atonal: homologue; orthologue; atonal-associated protein; deafness;
XX	hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KM	cellular proliferation; cerebellar granule neuron; gene therapy;
KM	mechanoreceptive cell growth; auditory; osteopathic; cyostatic;
XX	transgenic animal; 89.
XX	
OS	Mus musculus.
XX	
FN	WO200073764-A2.
XX	
PD	07-DEC-2000.
XX	
PF	01-JUN-2000; 2000WO-US15410.
XX	
PR	01-JUN-1999; 99US-0137060.
PR	19-JAN-2000; 2000US-0176993.
XX	
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.
XX	
PL	Zoghbi HY, Belien H, Birmingham N, Haasan B, Ben-Arie N,
XX	
DR	WPI; 2001-032190/04.
DR	P-PSDB; AAB60350.
XX	
PT	Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT	any of its homologs or orthologs, for the treatment of e.g. deafness,
XX	osteoarthritis and abnormal cell proliferation -
XX	
PS	Disclosure; Page -; 142pp; English.
XX	
CC	The invention relates to the use of atonal-associated nucleic acid or
CC	amino acid sequence, or any of its homologues or orthologues as
CC	therapeutic agents for the treatment of deafness, partial hearing loss,
CC	vestibular effects due to damage or loss of inner hair cells,
CC	osteoarthritis and abnormal cell proliferation. The invention also
CC	encompasses methods of screening for compounds which affect the
CC	expression of an atonal-associated nucleic acid sequence in an animal,
CC	and a transgenic animal in which an allele of a native atonal-associated
CC	gene is replaced by a heterologous nucleic acid sequence, thus
CC	inactivating the atonal-associated allele. The nucleic acids or proteins
CC	may be used in a method of treating an animal for hearing impairment,
CC	joint disease, balance disorders, abnormal cell proliferation, or other
CC	disease related to loss of a functional atonal-associated nucleic acid or
CC	protein. They may particularly be used to treat an animal with a
CC	deficiency in cerebellar granule neurons or their precursors, and may
CC	also be used in promoting mechanoreceptive cell growth and generating
CC	hair cells. The present sequence represents an atonal-associated nucleic
CC	acid sequence referred to in the invention.
CC	Note: The present sequence is not shown in the specification, but
CC	was obtained from GenBank.
XX	
XX	
SQ	Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
XX	
Query Match	56.2%; Score 819.8; DB 22; Length 5567;
Best Local Similarity	87.1%; Pred. No. 7.5e-213;
Matches 997; Conservative	0; Mismatches 102; Indels 46; Gaps 7;
OY	4 GGTACGAGAGAGAGATGCCCTGGGGCCCCCGTTGCTGATGTGGCCCGTGGACACAGCACA 63
DB	4424 GGGCTGCGGTGGCGAGGCCCTGGGGCCCCCTCGCTGATGTGGCCCGTGGTGGAGCACA 4483
OY	64 GCCCGGAGAGAGACGCTCTGTGTC - GGGCGAGAGAGATTAAGCTGCCAGGGGACACAG 122
DB	4484 GCCCGGAGAGACGCTCTGTGCGGGGGGCGAGAGATTAAGCTGCCAGGGGACACAG 4544
OY	123 ATT---AGCAGCTCAAAATGCTCTGTGGGTTCACCACTGA - CAGAGCGAGAGACC 177
DB	4544 ACTTCATGCAAGCTCAAAATCCCTGTGGGTTCATCACTGCAGCAGTGGTGAATACT 4603
OY	178 CCTCGAGGCTCTTTGCTGCGCTCCAGAGCAATTTATCTCAGAGAGGGCGGCTGCAGCT 237
DB	4604 CTTGCGAGCTTTTCTACGACTTTCAGAGCAATTTATCTCAGAGAGGGCGGCTGCAGTT 4663

QY	238	CAGCAAAA	CTTCGAGGAGGAGGAGGGGTTAGCTATCCACCGGTGCTGACCTGACC	296
Db	4664	TAGCAAC	CTTCAGAGGAGGAGGAGGAGGCTGAGCTATTCACCTGCTGACCTGACC	4723
QY	297	-----	-----ACCGCAGCTCTCTGTTCTTTTGTAGCCCGAGTA	330
Db	4724	TATCACT	GTGCTGTTCATGACTGACCTGTCTCTATCTTTTGTAGTGGGAG-A	4782
QY	331	ACTAGGT	TAACTTTTAGGAACCTCCAAAGGGTATGAAGAGGGAGT-----GGTGG	380
Db	4783	ACTAGGT	TAACTTCGAAACCTCCAAAGGGATGAGGGGGCGCGGGGTGTGTGGG	4842
QY	381	GCGA	ACTTACTCCCGGTGAGTGCCTCTAATCTAGAGCTGTCAACCC--CCTTC	438
Db	4843	GGAT	CTCTGTGCTCCCGGTGAGTGCCTCTAATCTAGAGGCTGTGACACACACCTTC	4902
QY	439	CATTTT	TTCCTCACTGAGATGGCGCTCATCCCTTGGATGCGCCACCATCCAACTGT	498
Db	4903	CATTTT	TTCCTCACTGAGATGGCGCTCATCCCTTGGATGCGCTCATCCATCCAACTGT	4962
QY	499	CCCAAG	AACCCAGAACCTTTTCCCGAGCTCTGGACACAGAACTGTCACTTCAATT	558
Db	4963	CCCAAG	AACCAACAACCTTTTCCCGAGCTCTGGACACAGAACTGTCACTTCAATT	5022
QY	559	CCACCC	CACTAGGCCACTCTGATCCGAGGGCTGTCCGAGCAAGACAGTGTACT	618
Db	5023	CAACCC	CACTAGGCCACTCTTCACTTCTAGGAGCTGTCCGAGCAAGAACTGTGTACT	5082
QY	619	GCCGAG	GACATCGAGGAAGCTCCGTGCGCGCGCGAGAGGGCGCAACAGGCCAAGACG	678
Db	5083	GCCGAG	GAGGACCTCGAGGAAGCTCCGTGCGCGCGCGAGAGGGCGCAACAGGCCAAGACG	5142
QY	679	AGTTG	CACTAGCAAGCAAGCAAGCAAGCGGCGCAAGAGCCCAAGACCGGGAGGGCA	738
Db	5143	AGTTG	CACTAGCACTAGCAAAACAGCAAGAAAGCGGCGCAAGAGCCCAATGTGGAGCGCA	5202
QY	739	ACCG	ATSCAACAACCTTAACCTCCGCGCTGATGTGCGTGCCTGTCTGCGCCACTTCC	798
Db	5203	ATCG	ATSCAACAACCTTAACCTCCGCGCGCTGATGTGCGTGCCTGTCTGCGCCACTTCC	5262
QY	799	CGGAT	GACGCGAACTTACAAAGATCGAGACCTCTGCGCTTGCCTCAACTACATTTGGG	858
Db	5263	CGGAT	GACGCGCGCAACTTACAAAGATCGAGACCTCTGCGCTTGCCTCAACTACATTTGGG	5322
QY	859	CAC	TGACCTCAACCTGTGGCACTACCGGACCAACACTTCTAAGCGGCCGAGCCCTGTGC	918
Db	5323	CAC	TGACCTCAACCTGTGGCACTACCGGACCAACACTTCTAAGCGGAGCCCTGTGC	5382
QY	919	CCTG	TGGGAGCTGTGGAAAGCCCGGGAAGGGGCTCTCAGCGGCGCATGTGGGGCTCTACT	978
Db	5383	CCTG	TGGGAGAGCTGTGGGAGCCCGGAGGTGGCTCCAAAGGGGACTGTGGGGCTCTACT	5442
QY	979	CCCC	AGTTTCCCAAGCTGTGAGCCTGAGCCCCACAGCCTCATTTGAGAGAGTTCCCTGGCC	1038
Db	5443	CCCC	AGTCTCCCAAGCGGGTAACTGTAGCCCAAGCGCTCATTTGAGAGAAATTCCTGTGCC	5502
QY	1039	TGC	AGGTGCCCACTCCCAATCTGTCTGTCTCCGGGGCACCTGTGTCTTCAACATTTCT	1098
Db	5503	TGC	AGGTGCCCACTCCCAATCTGTGTCTGTCTCCGGGAGCACTGTGTCTTCAACATTTCT	5562
QY	1099	TGTG	A1103	
Db	5563	TGTG	A5567	
RESULT 5				
AAF27266				
ID AAF27266 standard; cDNA; 861 BP.				
XX AAF27266;				
DT 24-APR-2001 (first entry)				





XX 19-APR-2002; 2002WO-EP04362.  
 PF basic helix-loop-helix transcription factor neurogenin 3 (ngn3),  
 XX a gene which is required for the specification of the early  
 PR 19-APR-2001; 2001US-284531P.  
 XX (DEVE-) DEVELOPMENTAL BIOLOGISCHE FORSCH.  
 PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.  
 XX Mobus AM, St-Onge L, Blyszczuk P, Hoffmann U;  
 XX WPI; 2003-075629/07.  
 DR Differentiating stem cells into insulin-producing cells useful for  
 PT treating pancreatic diseases, by culturing stem cells in suitable  
 PT medium and activating gene involved in beta-cell differentiation -  
 XX  
 PS Disclosure; Page 58-59; 62pp; English.  
 XX  
 CC The present sequence is that of cDNA encoding the murine  
 CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),  
 CC a gene which is required for the specification of the early  
 CC endocrine precursor in the pancreatic epithelium and which is  
 CC down-regulated once endocrine differentiation begins. The invention  
 CC provides a claimed method for differentiating stem cells (especially  
 CC embryonic, adult or somatic stem cells and primordial germ cells)  
 CC into insulin-producing cells. This involves culturing stem cells in  
 CC a suitable medium and activating at least one gene involved in  
 CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,  
 CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the  
 CC delivery of the gene into stem cells using a viral delivery  
 CC system, or the delivery of a protein product of the gene into stem  
 CC cells. The insulin-producing cells can be transplanted into  
 CC animals or human for treatment of pancreatic diseases, metabolic  
 CC syndrome and metabolic disorders with impaired glucose levels such  
 CC as diabetes, hyperglycaemia and impaired glucose tolerance  
 CC (claimed). The cells can also be used to identify compounds which  
 CC stimulate beta-cell differentiation, insulin secretion or glucose  
 CC responsiveness. Differentiated beta-cells can be used to study the  
 CC toxic and other effects of exogenous compounds on beta-cell  
 CC function. In an example from the invention, Pax6 cDNA was inserted  
 CC into expression vector PACCMV.pLpA under the control of the  
 CC cytomagalovirus promoter.  
 CC  
 CC Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;  
 XX  
 SQ  
 Query Match 46.7%; Score 682.4; DB 25; Length 861;  
 Best Local Similarity 89.7%; Pred. No. 1e-175;  
 Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;  
 QY 312 TTCTTTGAGCCCGGAGTAACTTAACTTCAAGGAGGAGGAGG 371  
 DB 2 TTCTTTTAACTGCGGAG- AACTAGTAACTTGGAAATCTCAAGGATGATAGGAGG 60  
 QY 372 AGTGGT-----GGGCGTACTAGTCCGCGTGAAGTGAAGTCTTAAGTCAGAGA 421  
 DB 61 CGCGGGGAGTGTGTGGGGGATCTCTGTGTCCTCCGCGAGTGAAGTCTTAAGTCAGAGG 120  
 QY 422 CTG--TCACACCCCTTCCATTTTTCCTCAAGTCTGAGATGAGGCTCTTCTTGAAT 479  
 DB 121 CTGGACACACACACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 180  
 QY 480 GCGCCACATCCAGTGTCCCAAGAGCCGAGAACCTTTCCGAGAGCTCCGAGCAGC 539  
 DB 181 GCGCTACATCCAGTGTCCCAAGAGCCGAGAACCTTTCCGAGAGCTCCGAGCAGC 240  
 QY 540 GAAGTGTCTAGTCTCAATTCACCCACCTAGCCCTCTCTGTAACGAGGAGTGTCTCC 599  
 DB 241 GAAGTGTCTAGTCTCAATTCACCCACCTAGCCCTCTCTGTAACGAGGAGTGTCTCC 300  
 QY 600 GAAGGAGAGAGAGTGTCTGAGGAGGAGATGAGGAGGAGTCTGAGGAGGAGGAGG 659  
 DB 301 GAAGGAGAGAGTGTGAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360

QY 660 CGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACAGCAGCAGAGAGCCGCGCAAGAG 719  
 DB 361 CGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACAGCAGCAGAGAGCCGCGCAAGAG 420  
 QY 720 GCCAAGCAGCGGGAGCGCAACCGCATGACAACTTAACTCGGCTGGATGCGTGGC 779  
 DB 421 GCCAATGATCGGGAGCGCAATGATGACAACTTAACTCGGCTGGATGCGTGGC 480  
 QY 780 GGTGTCTGCGCCACCTTCCGAGATGAGCCCAACTTAAAGATGAGAGCCCTGCTTC 839  
 DB 481 GGTGTCTGCGCCACCTTCCGAGATGAGCCCAACTTAAAGATGAGAGCCCTGCTTC 540  
 QY 840 GCCCAACATCAATTTGGGACCTGACTGACGCTGCGCATAGCGGACACAGCTTCTAC 899  
 DB 541 GCCCAACATCAATTTGGGACCTGACTGACGCTGCGCATAGCGGACACAGCTTCTAT 600  
 QY 900 GGGCCCGAGCCCTCTGTGCTGTGGGAGCTGGGAGAGCCCGGAGGGGGCTCCAGCGC 959  
 DB 601 GGGCCCGAGCCCTCTGTGCTGTGGGAGCTGGGAGAGCCCGGAGGGGGCTCCAGCGC 660  
 QY 960 GACTGGGCTCTATCTACTCTCCAGTTTCCCAAGCTGGTACCTGAGCCCAAGCCTCA 1019  
 DB 661 GACTGGGCTCTATCTACTCTCCAGTTTCCCAAGCTGGTACCTGAGCCCAAGCCTCA 720  
 QY 1020 TTGAGAGAGTTCTCTGCTGCTGAGGTCGAGCTCCCATCTGTCTCTCCGGGACCC 1079  
 DB 721 TTGAGAGAGTTCTCTGCTGCTGAGGTCGAGCTCCCATCTGTCTCTCCGGGACCC 780  
 QY 1080 CTGCTGTCTCTGAGCTTCTGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1139  
 DB 781 CTGCTGTCTCTGAGCTTCTGTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 840  
 QY 1140 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1159  
 DB 841 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 860  
 RESULT 7  
 ID AAV27050 standard; cDNA; 804 BP.  
 XX AAV27050;  
 AC AAV27050;  
 DT 17-SEP-1998 (first entry)  
 XX  
 DE Mouse neurogenin 3 gene.  
 XX ds; Mouse; neurogenin; expression vector; recombinant protein;  
 KW antibody; neurogenesis.  
 XX  
 OS Mus sp.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 160..804  
 FT /\*tag= a  
 FT /product= "Mouse neurogenin 3"  
 XX  
 PN W09813491-A2.  
 XX  
 PD 02-APR-1998.  
 XX  
 PF 24-SEP-1997; 97WO-US17048.  
 XX  
 PR 17-SEP-1997; 97US-0932411.  
 PR 27-SEP-1996; 96US-0722570.  
 PR 12-NOV-1996; 96US-0030864.  
 PR 19-DEC-1996; 96US-0772009.  
 XX  
 PA (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 XX  
 PI Anderson DJ, Ma Q, Sommer L;  
 XX WPI; 1998-230702/20.

DR P-PSDB; AAW54947.  
 XX Mouse neurogenesis, useful in neurogenesis - and recombinant nucleic  
 PT acids and proteins derived from rat and Xenopus  
 XX  
 XX PS Disclosure; Fig 9; 106pp; English.  
 XX  
 CC The Mouse neurogenin 3 is one of several neurogenin proteins discussed  
 CC in the present invention. The neurogenin nucleic acids can be expressed  
 CC in a host cell, transformed using an expression vector, to produce  
 CC recombinant proteins. The proteins and the antibodies raised against  
 CC the proteins are useful in the study of neurogenesis.  
 XX  
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Query Match 44.7%; Score 652; DB 19; Length 804;  
 Best Local Similarity 90.9%; Pred. No. 1.9e-167;  
 Matches 731; Conservative 0; Mismatches 60; Indels 13; Gaps 3;

312 TTCTTTGAGCCCGAGTAAGTAACATTAGGAACCTCCAAAGGTAGAGAGGG 371  
 2 TTCTTTGAGTGGAG-ACCTAGTAACATTCCGAACTCCAAAGGTAGAGGG 60  
 372 AGTGGT-----GGGCGTACTCTAGTCCCGGTGAGTGAACCTTAAGTCAAGA 421  
 61 CGCGGGGGGTGTGTGGGGGATGACTGTGTCCTCCGCTGCACTGACCTTAAGTCAAGG 120  
 422 CTG--TCACACCCCTCCATTTTTCCTCCAACTTCAGATGCGGCTCATCCCTTGAT 479  
 121 CTGGCACACACACACCTTCATTTTTCCTCCAACTTCAGATGCGGCTCATCCCTTGAT 180  
 480 GCGCCACCATCAAGTGTCCCAAGAGACCCGAAACCTTCCCGAGGCTCGGACAC 539  
 181 GCGCTCACCATTCAAGTGTCCCAAGAGACCCGAAACCTTCCCGAGGCTCGGACAC 240  
 540 GAAGTGTCTAGTTCATTTTCAACCCGACCTAGCCCTCTGTACCGAGGACTGCTCC 599  
 241 GAAGTGTCTAGTTCATTTTCAACCCGACCTAGCCCTCTGTACCGAGGACTGCTCC 300  
 600 GAAGCAGAGAGAGTGTCTCCGAGGAGATCGAAGAACTCCGTGCGGGCGGAGGG 659  
 301 GAAGCAGAGAGAGTGTCTCCGAGGAGATCGAAGAACTCCGTGCGGGCGGAGGG 360  
 660 GCGACAGAGCCCAAGAGAGAGTGTGCACTGAGCAAGACGAGAGCGGCGCAAGAG 719  
 361 GCGACAGAGCCCAAGAGAGAGTGTGCACTGAGCAAGACGAGAGCGGCGCAAGAG 420  
 720 GCGAACGAGCCGAGAGAGCGAATGCAACCTTAAGTCCGCGCTGAGTCCGCTGCG 779  
 421 GCGAATGATCGGAGAGCGAATGCAACCTTAAGTCCGCGCTGAGTCCGCTGCG 480  
 780 GGTGTCTCTGCGCACTTCCCGAGTATGCGCAAACTTAAGTAAGTCAAGACCTGCGCTTC 839  
 481 GGTGTCTCTGCGCACTTCCCGAGTATGCGCAAACTTAAGTAAGTCAAGACCTGCGCTTC 540  
 840 GCGCAACTATATTTGGGCACTGATGAGAGTGTGCGAGTGTGCGGAGCAAGCTTCTAC 899  
 541 GCGCAACTATATTTGGGCACTGATGAGAGTGTGCGAGTGTGCGGAGCAAGCTTCTAC 600  
 900 GCGCCCGAGAGCCCTGTGCTGTGCGGAGTGTGAGAGCCGCGGAGGTGCTCAAGCGCG 959  
 601 GCGCCCGAGAGCCCTGTGCTGTGCGGAGTGTGAGAGCCGCGGAGGTGCTCAAGCGCG 660  
 960 GACTGGGGCTATCTACTCTCCAGTTTCCAAAGTGTGAGCTGAGCCCAAGCTCA 1019  
 661 GACTGGGGCTATCTACTCTCCAGTTTCCAAAGTGTGAGCTGAGCCCAAGCTCA 720  
 1020 TTGGAGGAATTCCTGCGCTGAGAGTCCCAAGTCTCCATCTGTGTCTCCCGGGACAC 1079  
 721 TTGGAGGAATTCCTGCGCTGAGAGTCCCAAGTCTCCATCTGTGTCTCCCGGGACAC 780  
 1080 CTGTGTCTCAGACTTCTTGTA 1103

DB 781 CTGTGTCTCAGACTTCTTGTA 804

RESULT 8  
 ID AAZ51981 standard; DNA; 804 BP.  
 AC AAZ51981;  
 DT 04-JUL-2000 (first entry)  
 DE Murine neurogenin-3 (NGN3) nucleic acid sequence.  
 XX  
 XX Murine neurogenin-3; non-neuronal cell; NNC; neurogenesis;  
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;  
 KW neural differentiation; transplantation; neuronal dysfunction;  
 KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;  
 KW neuroprotective; nocitropic; anticonvulsant; antiParkinsonian; vulnerary;  
 KW cerebrioprotective; immunosuppressant; antiinfectious; ds.  
 XX  
 XX Mus sp.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 160..804  
 FT /tag=a  
 FT /product="Murine neurogenin-3 protein"  
 XX  
 XX MO200009676-A2.  
 XX  
 XX 24-FEB-2000.  
 XX  
 XX 13-AUG-1999; 99MO-US18525.  
 XX  
 XX 14-AUG-1998; 98US-0096630.  
 XX  
 XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
 PA  
 PI Anderson DJ, Lo L;  
 XX  
 XX MPI: 2000-256250/22.  
 DR P-PSDB; AAY70570.  
 DR  
 XX  
 XX Inducing non-neuronal cells to differentiate into neurons and for  
 PT non-neuronal cells to express a neuronal subtype-specific marker,  
 PT comprising contacting the non-neuronal cells with a vector containing  
 PT neurogenin nucleic acid -  
 XX  
 XX Claim 1; Fig 1J; 76pp; English.  
 PS  
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to  
 CC differentiate into neurons and for NNCs to express a neuronal subtype  
 CC -specific marker. Transformed host cells are used as sources of neuronal  
 CC and other growth factors; in culture for screening compounds that  
 CC modulate neural differentiation or as sources of recombinantly produced  
 CC neurogenins and Phox2a proteins for use in transplantation. The cells  
 CC also have a variety of in vivo uses, e.g., for transplantation at sites of  
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to  
 CC optical or auditory nerve damage, brain or spinal cord injuries, and  
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present  
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.  
 CC NNCs differentiate into neurons through the recombinant expression of a  
 CC transcription factor that induces a core program of neurogenesis. Forced  
 CC expression of murine NGN3 can elicit expression of at least some neuronal  
 CC phenotypic markers even in NNCs.  
 CC  
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Query Match 44.7%; Score 652; DB 21; Length 804;  
 Best Local Similarity 90.9%; Pred. No. 1.9e-167;  
 Matches 731; Conservative 0; Mismatches 60; Indels 13; Gaps 3;

312 TTCTTTGAGCCCGAGTAAGTAACATTAGGAACCTCCAAAGGTAGAGAGGG 371

```

Db      2 TTCTTTGATGCGGAG-AAC TAGGTAACAATTGCGAAAATCTCCAAAGGGTGATGAGGGG 60
Qy      372 AGTGGT-----GGCGCTACTCTAGTCCCGCTGAGTGAAGCTTAAGTCAGAGA 421
Db      61 GCGCGGGGTGTGTGTGGGGATTAATCTGTGTCCCGGTGCACTGACTTAAGTCAGAG 120
Qy      422 CTG--TCACACCCCTTTCATTTTTCCTCAACCTCAGAGATGGCGCTCATCTTGGAT 479
Db      121 CTGGCACACACACCTTTCATTTTTCCTCAACCTCAGAGATGGCGCTCATCTTGGAT 180
Qy      480 GCGCCACCATTCAGAGTGTCCCAAGAGACCAAGAACCTTTCCTGGAGGCTTCGACAC 539
Db      181 GCGCTACCATTCAGAGTGTCCCAAGAGACCAAGAACCTTTCCTGGAGGCTTCGACAC 240
Qy      540 GAAGTGTCTAGTTCATTTCCACCCCTAGAGCCCTCAGTCTGATCCAGAGGACTGTCC 599
Db      241 GAAGTGTCTAGTTCATTTCCACCCCTAGAGCCCTCAGTCTGATCCAGAGGACTGTCC 300
Qy      600 GAAGCAGAAGAGTGTGACTGCGGAGGAGACATCAGAGAAAGTCTGCGCGCGCGAGAG 659
Db      301 GAAGCAGAAGAGTGTGACTGCGGAGGAGACATCAGAGAAAGTCTGCGCGCGCGAGAG 360
Qy      660 CGCAACAGAGCCCAAGAGAGGAGTGTGACTGAGCAAGACCGCAAGCGCGCGCAAGAG 719
Db      361 CGCAACAGAGCCCAAGAGAGGAGTGTGACTGAGCAAGACCGCAAGCGCGCGCAAGAG 420
Qy      720 GCCAAGCAGCGGAGAGCGCAACCGCATGCAAACTTAATCTCGCGCTGATGCGCTGGCG 779
Db      421 GCCAATGATCGGAGCGCAATGCAATGCAAACTTAATCTCGCGCTGATGCGCTGGCG 480
Qy      780 GGTGTCTGCGCCACCTTCCCGATGACGCCAACTTAAGAGATGAGAACCTGCGCTTC 839
Db      481 GGTGTCTGCGCCACCTTCCCGATGACGCCAACTTAAGAGATGAGAACCTGCGCTTC 540
Qy      840 GCCCAACATCATTTGGGCACTGACTAGAGCGTGGCGATGCGGACCAAGCTTTTAC 899
Db      541 GCCCAACATCATTTGGGCACTGACTAGAGCGTGGCGATGCGGACCAAGCTTTTAT 600
Qy      900 GGGCCGAGCGCCCTGTGCTGCGGAGAGTGGGAGCGCGGAGGGGGCTCCAGCGGG 959
Db      601 GGGCCGAGCGCCCTGTGCTGCGGAGAGTGGGAGCGCGGAGGGGGCTCCAGCGGG 660
Qy      960 GACTGGGCTCTATCTACTCTCCCAAGTGTGAGCTGAGGCCCAAGCAAGCTTCA 1019
Db      661 GACTGGGCTCTATCTACTCTCCCAAGTGTGAGCTGAGGCCCAAGCAAGCTTCA 720
Qy      1020 TTGAGAGATTCCTGCGCTGAGAGTGGCCAGCTTCCCATCTGTCTGCTCCGGGCA 1079
Db      721 TTGAGAGATTCCTGCGCTGAGAGTGGCCAGCTTCCCATCTGTCTGCTCCGGGCA 780
Qy      1080 CTGGTCTCTGAGACTTCTTGA 1103
Db      781 CTGGTCTCTGAGACTTCTTGA 804

```

RESULT 9  
AAC61089  
ID AAC61089 standard; DNA; 5340 BP.

AAC61089;  
AAC61089;  
05-FEB-2001 (first entry)  
Human neurogenin 3 (Ngn3) genomic DNA sequence.  
Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;  
islet cell precursor identification; diabetes mellitus; human; ds.  
Homo sapiens.  
Key Location/Qualifiers  
FH 3022..3666  
FT CDS  
/\*tag= a

```

FT      /product= "Ngn3"
FT      /note= "Neurogenin 3"
PN      MO200059936-A1.
XX      12-OCT-2000.
XX      28-MAR-2000; 2000MO-US08436.
XX      06-APR-1999; 99US-0128180.
PR      (RBC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI; 2000-664989/64.
DR      P-PSDB; AAY85617.
XX      Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT      them, useful for diagnosis, prevention and treatment of diabetes
PT      mellitus and to identify individuals at risk of diabetes -
XX      Claim 6; Page 46-48; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC      protein AAY85617. The Ngn3 gene is located at chromosome position
CC      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC      protein sequences, and includes an antibody recognising the Ngn3 protein.
CC      Also included in the invention is a method for identifying an islet cell
CC      precursor, the method involves analysing a cell for the expression of the
CC      Ngn3 gene product, where detection of the product is indicative of an
CC      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC      reagent for detecting (in a subject) a predisposition to a defect in
CC      pancreatic islet cell function or formation associated with a defect in
CC      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC      precursor cells expressing Ngn3, and to alter cellular differentiation in
CC      culture in vivo to produce new beta-cells to treat patients with diabetes
XX      mellitus.
SQ      Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

```

Query Match 35.4%; Score 517.4; DB 21; Length 5340;  
Best Local Similarity 67.4%; Pred. No. 1.8e-130;  
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

```

Qy      20 GTCCCTGGGCCCCCTTGTCTATTTGCCCCGTGGGACAGAGCCCGGAGAGCAGCT 79
Db      2535 GGGCCAGAGGCCCCGGCGCTGATTGGCCGATGGCGCGGAGAGCCGAGAGCAGCT 2594
Qy      80 CCTGATCGGCGCAGAGCAGATPAAAGCGTGCAGAGGAGACACAGATTAGCACTCAGAA 139
Db      2595 CTTGGCCCGGCGAAGCAGATPAAAGCGTGCAGAGGAGACACAGATTAGCACTCAGAA 2654
Qy      140 CCTCTGTGGTCTCAACTGC-----ACAGAGCCGAGAGACCCCTC 181
Db      2655 TCCCTGGGCTTCAACCGCGGCTCGAGAGAGCGTGAAGAGGCTTCGAGCCCAT 2714
Qy      182 CGAGCTTTTGTGCTCTCCAGAGCAATTACT-CCAGGAGAGGGCGCTGCACTAG 240
Db      2715 CTCTCTTCTTTTCTCTTTGGGGCTGGGCAACTCCAGAGGGGGCGCTGCACTAG 2774
Qy      241 CAAAATTGGAAGGAGAG-----AGGGGTTCAGCTATCCAGCCCTGCT 285
Db      2775 CTGAACCTTGGAGCAGACAGAGCCGCTGAGCTCCCAAGCGCCCTGCTCATAGCTTTC 2834
Qy      286 TGACTGTGACCAACCGCAGCTCTCTGTCTTTTGAAGCCCGAGATPAACTAATTA 345
Db      2835 TATCTTTTGGCGCGATGAAGATTAATTGAGAGCTCCGAGGAGCGGAGGAGAA 2894
Qy      346 GGAAC-CTCCAAAGGATGAAGAGGAGTGGTGGCGTACTAGTCCCGCTGAGT 404
Db      2895 AGAGGATCTCTGACCCAGCGGGGCTGGAGAGATGAGTGTCTTTTGTCTCCACCT 2954

```

OY	405	GACCCCTTAAGCTCAGAGACTG-----TCACACCCCCCTTCCATTTTTTCCCAA	451
Db	2955	AGCCTCGGAATGCGGACTGCGCCGTGACAGACTAAACTTACCTTCCCTCTGACCCCG	3014
OY	452	CCTCAGATGAGGCGCTCATCCCTTGATGCGGCCCACTCCAAAGTGTCCCAAGACCCA	511
Db	3015	CGTATAGATTGAGCCTTCAACCTCGGGTGCGCCCATGTCCAAAGTGAACCCGTAACGGA	3074
OY	512	GCACCCCTTTCCCGGAGCCTTGGGACCAAGAGTGTCTCAAGTTCCAAATTCACCCACTAG	571
Db	3075	GCGGTCTTCCCCCAAGGCTTCGGAAGAGAGTGAACCTGCCCAAGTCCGCCCCCGCAG	3134
OY	572	CCCCACTCTCGATCCGAGGAACTGCTCCGAAAGCAAGAGGTGACTGCCGAGGGCAATC	631
Db	3135	CCCCACTGACACACGGGGAAATGTGCGACAGAGCGGAAGGGAGGCTGCGAGGGGGCCC	3199
OY	632	GAGAAAGCTCGTGCGCGCGCGGAGGGGCGCAACAGGCCCAAGACGAGTTGGCACTGAG	691
Db	3195	GAGGAAGTCTCGGGGACCGCGCGGGGGAGACGACCGGCTTAAGACGAGTTGGCACTAG	3254
OY	692	CAAGCAGCGACGAGCCGCGCGCAAGAGGCCAAGACCGGGAGCGCAACCGATGCACAA	751
Db	3255	CAAGCAGCGACGAGTTCGCGCAAAAGAGGCCAAGACCGCGAGCGCAATCGAATCAACA	3314
OY	752	CCTTAATCTCCGGCTGGAATGCGCTGCGGGGTCTCCTGCCACACTTCCCGGATGAGCCCA	811
Db	3315	CCTCAACTTCGGACTGGAAGCGCTTCCGGGTCTCTGCGCACTTCCACAGACGACGCA	3374
OY	812	ACTTACAAAGATCGAGACCCCTGCGCTTGCGCCCAACTATTGGCACTGACTCAAG	871
Db	3375	GCTACCAAGATCGAGACGCTGCGCTTGCGCCCAACTATCTGGGCGCTGACTCAAC	3433
OY	872	GCTGCGATAGCGGACCAACAGCTTCTACCGGCCCGAGCCCTGTGCTGTGGGAGCT	931
Db	3435	GCTGCGATAGCGGACCAACAGCTTGTACGCGCTGAGCGCGCGCGCACTGCGGGGA	3494
OY	932	GGAAGACCCCGGAGGGGGGCTCCAGCGGCGACTGCGGGCTTATATCTATCTCCCACTTCCCA	991
Db	3495	GCTGGGACAGCCCAAGCGGTTCCCGCGGGACTGGGGGATCCCTCTACTCCCAAGTCTCCA	3554
OY	1052	CTCCCATCTGTGTGCTTCCCGGGCAACCTGTGTGTTCACACTTCTGTGAAGGCCCA	1111
Db	3615	CTCTTCCGCGCTGCTTGAAGCCAGGACAGTCTGACTTCTCAGATTCTGTGAAGGACT	3674
OY	1112	AACAGAGCCCTGGGCGGTGGGCGCTTGACAGAAAGGAGGAGTCAAGCTGTGTGAATGG	1177
Db	3675	GTTCTGTCTCTGGGCTGTGGGTCTTAAGGTTAAGGAGAGGAGACCGGAGGCGGTAG	3734
OY	1172	AAGGTAGTGAGGACACTCGAGCATC	1196
Db	3735	AGGGTGGCCGACGCGCGCGGCCCTC	3759
RESULT 10			
AADD6871			
ID	AADD6871	standard; DNA; 5340 BP.	
XX	AADD6871;		
XX	27-JAN-2003	(first entry)	
DT			
XX			
DE		Human neurogenin 3 (Ngn3) gene.	
XX			
KW		Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;	
KW		type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;	
KW		islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q21.1-22.2;	
KW		gene; db.	
XX			
OS		Homo sapiens.	

XX	Key	location/Qualifiers
FH	CDS	3022..3666
FT		/tag=a
FT		/product= "Human Ngn3 protein"
PX		
PN		MO200274045-A2.
XX		
XX		26-SEP-2002.
PD		
XX		20-MAR-2002; 2002WO-US11166.
PF		
XX		20-MAR-2001; 2001US-0817360.
PR		
XX		(RBC ) UNIV CALIFORNIA.
PA		
XX		German MS, Lin J;
PI		
XX		WPI; 2002-759853/82.
DR		P-FSDB; AAE29277.
XX		
XX		producing a mammalian islet cell for treating diabetes mellitus
PT		comprises introducing into a mammalian cell a nucleic acid molecule
PR		encoding neuroendocrine basic helix-loop-helix transcription factor -
XX		
PS		Example 2; Page 87-88; 108pp; English.
CC		The invention relates to a method for producing a mammalian islet cell.
CC		The method comprising introducing into a mammalian cell a nucleic acid
CC		molecule encoding an islet transcription factor for expression of the
CC		islet transcription factor in the cell and for production of islet cell
CC		phenotype in the cell. The islet transcription factor is a neuroendocrine
CC		basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC		for treating type 2 diabetes mellitus and for replacing beta cells lost
CC		to autoimmune destruction in individuals with type 1 diabetes. The method
CC		is useful in cell therapy. The present sequence is human neurogenin 3
CC		(Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
XX		
SQ		Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;
Query Match	35.4%; Score 517.4; DB 24; Length 5340;	
Best Local Similarity	67.4%; Pred. No. 1.8e-130;	
Matches 826; Conservative	0; Mismatches 351; Indels 48; Gaps 5;	
OY	20	GTCCTGGGCCCGGTTCTGATTTGGCCCGTGGCAGACAGCACGCCCGGCAGGCACGCT 79
Db	2535	GAGCCCAAGGCGCCGGCGGTATTGGCCGGTGCGCGGAGCAGACCAGCGGCAGGCACGCT 259
OY	80	CCTGGTCCGGGCAAGCAGATAAAGCCGCGCAGGAGAACACAGATTACAGTCAAGAT 139
Db	2595	CCTGGCCCGGCGGAAGCAGATTAAGCGTCCCAAGGGGCAACAGATTGCTGCAGAAA 265
OY	140	CCCTTGSGGTCTACCACTGC-----ACAGAGCCGAGAGCCCCCTC 181
Db	2655	TCCCTGGGTCTACACCGCCGCGCCTCGAGAGAGCGTAGAGAGGCTTCGGAACCCATT 271
OY	182	CGAGCTCTTTGCTGCTCTCCAGAGCAAATTACT-CCAAGCGAGGGCGCTGCACTCAG 240
Db	2715	CTCTCTTCTTTTCTCTTTGGGGCTGGGGCAACTCCAGAGCGGGGCGCTGCACTCAG 277
OY	241	CAAACTTCGAAGGAGCAG-----AGGGGTTCAAGTATTCACCGCTGCT 285
Db	2775	CTGAACCTTGGGAGCAGCAAGAAGCCCGTAGCTCCCAAGCGCCCTGCTGCTCATTTGCTCTC 283
OY	286	TGACTCTGACACCGGAGCTCTGCTCTTTTGAGCCCGAGATTAAGTAAATTTVA 345
Db	2835	TATCTTTTGGCGCGGTAAAGTAATTTTGAAGCCTCCGAGGAGCGGCAAGGAGAA 289
OY	346	GGAAC-CTCCAAAGGTGAAAGAGGGAGTGGGTGGGCTACTAGTCCCGGTGAGT 404
Db	2895	AGAGGAGATCTCTGACCCAGCGGGGGCTGGAGAGATGCTGTTTTTTTCCACCT 295
OY	405	GACCTCTAAGTCAAGACTG-----TCACACCCCCTTCATTTTTCACAA 451



Db 117 CTTACCACTTTCCTCCCAACGACGGAACTCCAAATGAAACGTTAGCTTGGCCCA 58

Qy 845 CAATCACTTTGGGCACTGACTAGACGTCGCAATAGCGACACAGCTTTACG 900

Db 57 CAATTAATATTAACGTTAACTCAAGCGCTACCGGATTAACCAACCAACTTAAACG 2

## RESULT 12

ID ABQ49523 - standard; DNA; 592 BP

AC ABQ49523;

DT 12-JUL-2002 (first entry)  
yy

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.

KM Human, cytosine methylation; 5'-cpG-3'; uracil; cytosine; diagnosis;  
KM drug; side effect; cancer; central nervous system; cardiovascular;  
KM gastrointestinal; respiratory system; single nucleotide polymorphism  
KM SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2

PD 07-MAR-2002  
yy

01-SEP-2001; 2001WO-EPI0074

PR 01-SEP-2000; 2000DE-1043826

XX XX

PA (EPIC-) EPIGENOMICS AG

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ33410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other,

Query Match	15.7%	Score 229.6	DB 24	Length 592
Best Local Similarity	67.6%	Pred: No. 2e-52		
Matches 322; Conservative	0	Mismatches 154	Indels 0	Gaps 0

Qy	425	TCACACCCCCCTTCATTTTTTTCCAACTCAGATGCGGCTATCCCTTGGATGGCC	484
Db	116	TCAAATTACCCCTCCCTCTAACCCCGCGTAAATAACGCGCTCAACCCCTGAAATAGCC	175
Qy	485	CACCATTCCAAGTGTCCCAAGAGACCAGAACCTTTCCGGAGACCTTCGACCGAGAGT	544
Db	176	CATTAATCAAAATPAAACCGTAAACGAAAGCATCTCTTCCCAAAACCTCGAAAAAGAAAT	239
Qy	545	GCTCAGTTCCAAATTCACCCCACTPAGGCCCACTCTCGTACCGAGGACTGCTCGAAGC	604
Db	236	AACCTTACCCACGTCGCGCCCGCCCAACCCACTCTCGACAGAAAAAATTACGAAAAAC	299
Qy	605	AGAAGCAGGTGACTGCGAGGGGACATCGAGGAAGTCCGTGCGGCGCGGAGGGCCCA	664
Db	296	GAAAAAATAAACTACCGAAAAAACCCTCGAAACAGACGCGGAAAAACCAA	355
Qy	665	CAGGCCCAAGACGAGTTGGCATGTAGACAGACGAGAGCGGCGCAAGAAAGGCCAA	724
Db	356	CCGACTTAAAAAGAAATTAACATAACCAACAAACAGAGAAATCGACGAAAAAAACCAA	415
Qy	725	CGACCGGAGCGCAACCGCATCACAACCTTAACCTCCGCGTGGATGGCTGCGGGTGT	784
Db	416	CGAACCGGAACGGAATCGAATACACACTCAACTCGACATTAACGCGCTGAGCGATAT	475
Qy	785	CTGCGCAACCTTCGCGAATGACGCGAAATTCAAAAGTCGAAGCCTTGCGCTTCGGCCA	844
Db	476	CTAACCCACTTCTCCCAAGACGCGAAAATCAACCAAAATCGAAACGCTACGCTTGCCCA	535
Qy	845	CAACTCATTTGGGCGCATGACTCAGACGCTGGCGCATACGGACCAACAGCTTCTACG	900
Db	536	CAACTCATTTAAACGCTTAACCAACGCTAGCGCTAACGAAACCAACACTTATACG	591

### RESULT 13

ID ABQ49524 standard; DNA; 592 BP.

AC ABQ49524 ;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.

KM Human, cytosine methylation; 5'-CpG-3', uracil; cytosine; diagnosis; KM drug; side effect; cancer; central nervous system; cardiovascular; KM gastrointestinal; respiratory system; single nucleotide polymorphism; KM SNP; cell differentiation, ds.

OS Homo sapiens.

PN W0200218632-A2

PD 07-MAR-2002

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

XX

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. AB013410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;

Best Local Similarity 64.5%; Pred. No. 3.1e-42;

Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
OY 456 AGATGGGCGCTCATCCCTTGATGCGCCACCATCATCAAGTGTCCCAAGACCAGCA 515
    |||||
DB 147 AGAGATGAGTTTAAATTTTGGGTCGTTTATGTTTAAAGATTCGAGAGCGAGCG 206
OY 516 CCCCTTCCCGAGCCTCGAGACCAAGTGTCTCAATTCACCCCACTTACCCC 575
    |||||
DB 207 TTTTCTTTTAAAGTTTCCGAAAGACGAAGTGAATTTCTTTTACGTTTCGTTT 266
OY 576 ACTCCGTAACGAGGAGCTGCTCCGAGCAAGAGAGTGAAGTCCGAGGAGCAATCGAG 635
    |||||
DB 267 ATTCTATACGAGGGAATTTGCTAAGGCGGAAAGAGAGGAGTTGTCAGAGGTTTCGAG 326
OY 636 AAGCTCCGTGCGCGCGCGAGGAGCGCAACAGGCGCAAGAGGAGTGTGCACTGAGCAAG 695
    |||||
DB 327 AAGTTTCCGGTACGCGCGCGGAGACTGTACGCTTTAAGAGGAGTGTATGATGTAAG 386
OY 696 CAGCAGCAAGCCGCGCGCAAGAGGCGCAACGAGCGGAGCGCAACCGATGCACAACTT 755
    |||||
DB 387 TAGCAGCAGGAGTCCGCAAGAAAGGTTAAGATTCGAGCGTAAATGATATATTTT 446
OY 756 AACTCCGCGCTGAGTGGCTGCGCTGTCTCTCCCACTTCCCGAGTGAAGCCAACTT 815
    |||||
DB 447 AATTCGTAATGAGAGCTTTTGGCGGTGTCTTTGTTTATTTTGAAGACGCGCAAGTTT 506
OY 816 ACAAGATCGAGACCTGCGCTTCCGCCAACATCAATTTGGGCACTGACTGAGAGCTG 875
    |||||
DB 507 ATTAAAGATCGAGAGCTTGGCTTCTTTTAATTAATTAATTTGGCGTGAATTTAAACGTTG 566
OY 876 CGCATAGCGGAGCACAGCTTACG 900
    |||||
DB 567 CGTATAGCGGATTAATGTTTACG 591
    |||||
```

RESULT 14  
AB049525/C  
ID AB049525 standard; DNA; 592 BP.

XX AB049525;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
XX drug; side effect; cancer; central nervous system; cardiovascular;  
XX gastrointestinal; respiratory system; single nucleotide polymorphism;  
XX SNP; cell differentiation; ds.

OS Homo sapiens.  
XX WO200218632-A2.  
XX 07-MAR-2002.  
XX 01-SEP-2001; 2001WO-EPI10074.  
XX 01-SEP-2000; 2000DE-1043826.  
XX 05-SEP-2000; 2000DE-1044543.  
XX (EPIG-) EPIGNOMICS AG.  
XX Olek A, Piepenbrock C, Berlin K, Guetig D;  
XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful  
PT for diagnosis and prognosis, comprises selective hybridization of  
PT amplicons from chemically treated DNA

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of  
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
CC genomic sample of DNA. The sample is treated chemically to convert  
CC cytosine (C) but not methylated C, to uracil, then part of the genomic  
CC DNA that contains the target C is amplified to form a labeled amplicon.  
CC The amplicon is hybridized to two classes, each with at least one  
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
CC and the degree of hybridization to both classes is determined from the  
CC label on the amplicon. From the ratio of labels hybridized to the two  
CC classes of oligomers, the degree of methylation is calculated. The method  
CC is used: (i) for diagnosis and/or prognosis of side effects of  
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
CC of the central nervous, cardiovascular, gastrointestinal and respiratory  
CC systems etc., particularly by detecting mutations or single nucleotide  
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
CC types and for investigating cell differentiation. The method allows the  
CC methylation status of many C residues to be determined simultaneously.  
CC AB013410-AB054121 represent genomic DNA sequences used to illustrate the  
CC method for determining the degree of cytosine methylation described in  
CC the disclosure of the invention.

SO Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;

Best Local Similarity 64.5%; Pred. No. 3.1e-42;

Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
OY 456 AGATGGGCGCTCATCCCTTGATGCGCCACCATCAAGTGTCCCAAGACCAGCA 515
    |||||
DB 446 AGAGATGAGTTTAAATTTTGGGTCGTTTATGTTTAAAGATTCGAGAGCGAGCG 387
OY 516 CCCCTTCCCGAGCCTCGAGACCAAGTGTCTCAATTCACCCCACTTACCCC 575
    |||||
DB 386 TTTTCTTTTAAAGTTTCCGAAAGACGAAGTGAATTTCTTTTACGTTTCGTTT 327
OY 576 ACTCTGTAACGAGGAGCTGCTCCGAGCAAGAGAGTGAAGTCCGAGGAGCAATCGAG 635
    |||||
DB 326 AATTCGTAATGAGAGCTTTTGGCGGTGTCTTTGTTTATTTTGAAGACGCGCAAGTTT 267
OY 636 AAGCTCCGTGCGCGCGCGAGGAGCGCAACAGGCGCAAGGAGTGTGCACTGAGCAAG 695
    |||||
DB 266 AAGTTTCCGGTACGCGCGCGGAGAGCTGTACGCTTTAAGACGAGTTGTAATTAAGTAAG 207
OY 696 CAGCAGCAAGCCGCGCGCAAGAGGCGCAACGAGCGGAGCGCAACCGATGCACAACTT 755
    |||||
DB 206 TAGCAGCAGGAGTCCGCAAGAAAGGTTAAGATTCGAGCGTAAATGATATATTTT 147
OY 756 AACTCCGCGCTGAGTGGCTGCGCTGTCTCTCCCACTTCCCGAGTGAAGCCAACTT 815
    |||||
DB 146 AATTCGTAATGAGAGCTTTTGGCGGTGTCTTTTATTTTGAAGACGAGCAAGTTT 87
    |||||
```



QY 816 ACAAGATCGAGACCCCTGCGCTTCCGCCACACTACATTGGGCACTGACTCAGACGCTG 875  
DB 86 AATTAGATCGAGACCTTGCGCTTCTTTTAATTAATTTGGGCGTTGATTTAAACGTTG 27  
QY 876 CGCATAGCGGACCAACAGCTTCTTACG 900  
DB 26 CGTATAGCGGATTAATAGTTTGTACG 2

RESULT 15  
AAD46889  
ID AAD46889 standard; DNA; 714 BP.  
AC AAD46889;  
XX  
XX 27-JAN-2003 (first entry)  
DT  
XX  
XX Human neurogenin 1 (Ngn1) gene #2.  
DE  
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;  
KM type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;  
KM islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT 1..714  
FT CDS /tag= a  
FT /product= "Human Ngn1 protein"  
XX  
XX WO200274045-A2.  
XX  
XX 26-SEP-2002.  
XX  
XX 20-MAR-2002; 2002WO-US11166.  
XX  
XX 20-MAR-2001; 2001US-0817360.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX German MS, Lin J;  
PI  
XX  
XX WPI: 2002-759853/82.  
XX P-PSDB; AAE29280.  
XX  
XX  
XX Producing a mammalian islet cell for treating diabetes mellitus  
XX comprises introducing into a mammalian cell a nucleic acid molecule  
XX encoding neuroendocrine basic helix-loop-helix transcription factor -  
XX  
XX Disclosure; Page 94; 108pp; English.  
XX  
XX The invention relates to a method for producing a mammalian islet cell.  
XX The method comprising introducing into a mammalian cell a nucleic acid  
XX molecule encoding an islet transcription factor for expression of the  
XX islet transcription factor in the cell and for production of islet cell  
XX phenotype in the cell. The islet transcription factor is a neuroendocrine  
XX basic helix-loop-helix (bHLH) transcription factor. The method is useful  
XX for treating type 2 diabetes mellitus and for replacing beta cells lost  
XX to autoimmune destruction in individuals with type 1 diabetes. The method  
XX is useful in cell therapy. The present sequence is human neurogenin 1  
XX (Ngn1) gene.  
XX  
XX Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;  
SQ

Query Match 10.1%; Score 147.6; DB 24; Length 714;  
Best Local Similarity 71.2%; Pred. No. 4.7e-30;  
Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGCGGAGCGGCGGCAAGCGCCCAAGAGCGAGTTGGCACTGAGCAAGCGGAGCAAG 706  
DB 216 GCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 275

QY 707 CCGGCGCAAGAGCGCAACGACCGGAGCGGACCGGATGCAACCTTAATCCGCGCT 766  
DB 276 CCGGCGGCTCAAGCGCAACGATCGCGAGCGGCAACCGGATGCAACCTTAAGCGGCGCT 335  
QY 767 GGAATGCGGCTGCGGCGGTCTCTGCGCCACTTCCCGGATGAGCGCCAACTTAAGATGGA 826  
DB 336 GGAAGCACTGCGGAGCGGCTGCGCTGCGCTGCGGAGCAACCAAGCTCACCAAAATGGA 395  
QY 827 GACCGTGGCTTGGCCCAACACTACATTGGGCACTGACTCAGACGCTGCGCATAGCGGA 886  
DB 396 GACGCTGCGCTTCCGCTACAACTACATCTGAGGCTCTGGCGGAGACACTGCGGCTGCGGA 455  
QY 887 CCACAGCTTCTACGCGCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 920  
DB 456 TCAAGGCGTGGCCGAGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 489

Search completed: January 26, 2004, 19:39:01  
Job time : 430 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 26, 2004, 18:48:21 ; Search time 98 Seconds  
(without alignments)  
6575.708 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcaggtacgagagagagagcag.....agagtacctaaccagtcgt 1460

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents\_NA:\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTOS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	652	44.7	804	4 US-08-932-411A-19	Sequence 19, Appl
2	147.6	10.1	1268	1 US-08-910-973-12	Sequence 12, Appl
3	147.6	10.1	1268	4 US-09-499-227-12	Sequence 12, Appl
4	145.2	9.9	1527	4 US-08-722-570-12	Sequence 12, Appl
5	145.2	9.9	1527	4 US-08-932-411A-12	Sequence 12, Appl
6	140.4	9.6	738	4 US-08-722-570-13	Sequence 13, Appl
7	140.4	9.6	738	4 US-08-932-411A-13	Sequence 13, Appl
8	140.4	9.6	1333	4 US-08-910-973-21	Sequence 21, Appl
9	140.4	9.6	1333	4 US-09-499-227-21	Sequence 21, Appl
10	139.8	9.6	1385	4 US-08-932-411A-17	Sequence 17, Appl
11	121.8	8.3	310	1 US-08-552-142A-12	Sequence 12, Appl
12	115.6	7.9	1312	4 US-08-722-570-14	Sequence 14, Appl
13	115.6	7.9	1312	4 US-08-932-411A-15	Sequence 15, Appl
14	109.8	7.5	1277	4 US-08-722-570-15	Sequence 15, Appl
15	109.8	7.5	1277	4 US-08-932-411A-16	Sequence 16, Appl
16	96	6.6	1352	1 US-08-552-142A-10	Sequence 10, Appl
17	96	6.6	1535	1 US-08-910-973-10	Sequence 10, Appl
18	96	6.6	1535	1 US-09-499-227-10	Sequence 10, Appl
19	96	6.6	1550	3 US-09-234-332-3	Sequence 3, Appl
20	93.2	6.4	1462	1 US-08-552-142A-16	Sequence 16, Appl
21	93.2	6.4	1951	4 US-08-910-973-16	Sequence 16, Appl
22	93.2	6.4	1951	4 US-09-499-227-16	Sequence 16, Appl
23	92.4	6.3	2089	1 US-08-552-142A-1	Sequence 1, Appl
24	92.4	6.3	2089	1 US-08-910-973-1	Sequence 1, Appl
25	92.4	6.3	2089	4 US-09-499-227-1	Sequence 1, Appl
26	92.4	6.3	2089	5 PCT-US95-05741-1	Sequence 1, Appl
27	89	6.1	1560	1 US-08-552-142A-14	Sequence 14, Appl

28	89	6.1	1560	1 US-08-910-973-14	Sequence 14, Appl
29	89	6.1	1560	4 US-09-499-227-14	Sequence 14, Appl
30	89	6.1	2502	3 US-09-234-332-1	Sequence 3, Appl
31	85.8	5.9	1676	3 US-09-234-332-2	Sequence 2, Appl
32	83.4	5.7	524	1 US-08-552-142A-8	Sequence 8, Appl
33	83.4	5.7	524	1 US-08-910-973-8	Sequence 8, Appl
34	83.4	5.7	524	4 US-09-499-227-8	Sequence 8, Appl
35	83.4	5.7	524	5 PCT-US95-05741-8	Sequence 8, Appl
36	80	5.5	485	5 PCT-US95-05741-10	Sequence 10, Appl
37	73.6	5.0	1275	1 US-08-552-142A-3	Sequence 3, Appl
38	73.6	5.0	1275	1 US-08-910-973-3	Sequence 3, Appl
39	73.6	5.0	1275	1 US-09-499-227-3	Sequence 3, Appl
40	73.6	5.0	1275	5 PCT-US95-05741-3	Sequence 3, Appl
41	51.6	3.5	1635	3 US-09-234-332-4	Sequence 4, Appl
42	51.6	3.5	1635	4 US-09-702-705-1798	Sequence 1798, Ap
43	51.6	3.5	1635	4 US-09-736-457-1798	Sequence 1798, Ap
44	41	2.8	1785	5 PCT-US94-12912-1	Sequence 1, Appl
45	41	2.8	3636	1 US-07-753-520B-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-932-411A-19  
; Sequence 19, Application US/08932411A  
; Patent No. 6566496  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David J.  
; APPLICANT: Ma, Qifu  
; TITLE OF INVENTION: NEUROGENIN  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,411A  
; FILING DATE: 15-SEP-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/772,009  
; FILING DATE: 19-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/722,570  
; FILING DATE: 19-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 19:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 804 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 160..801  
; US-08-932-411A-19

Query Match 44.7%; Score 652; DB 4; Length 804;  
Best Local Similarity 90.9%; Pred. No. 4.6e-163;  
Matches 731; Conservative 0; Mismatches 60; Indels 13; Gaps 3;

```
QY 312 TTCTTTTGGAGCCGGAGTAATAGTAATCAATTAGAACTTCAAGAGGATAGAGAGG 371
DB 2 TTCTTTTGGAGGAG--AACTAGGTAACAAATTCGAAACCTCAAGAGGATAGAGAGG 60
QY 372 AGTGGT-----GGCGTACTCTAGTCCCGCGGAGTACCTCTAAGTCAAGAG 421
DB 61 GCGCGGGGATGTGTGGGGGTAATCTGTGTCTCCCGGTGCACTGTAAGTCAAGAG 120
QY 422 CTG--TCACACCCCTCCATTTTTCACACCTCAAGATGGGCTCATCTTGGAT 479
DB 121 CTGGCAACACACACCTTCCATTTTTCACACCTCAAGATGGGCTCATCTTGGAT 180
QY 480 GCGCCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCGCGAGCTCGAGCAG 539
DB 181 GCGCTCACCATCCAAAGTGTCCCAAGAGACCCAGCAACCTTTCGCGAGCTCGAGCAG 240
QY 540 GAAGTGTCTAGTTCATTCACCCCACTTACCCCACTCTGTACCGAGGATCTGCTC 599
DB 241 GAAGTGTCTAGTTCATTCACCCCACTTACCCCACTCTGTACCGAGGATCTGCTC 300
QY 600 GAAGAGAGAGAGTGTCTGCGAGGAGACATGAGAGAGCTCGTGGCGCGCGAGAGG 659
DB 301 GAAGAGAGAGTGTGTGATCTGCGAGGAGACATGAGAGAGCTCGTGGCGCGCGAGAGG 360
QY 660 CGCAACAGGCGCCAGAGCGAGTGTGCACTGAGCAAGCAGAGCGAGCGCGCAAGAG 719
DB 361 CGCAACAGGCGCCAGAGCGAGTGTGCACTGAGCAAGCAGAGCGAGCGCGCAAGAG 420
QY 720 GCGAAGAGAGCGGAGCGCAACCGCATGCAACCTTAATCTCGCGCTGATCGCTGCGC 779
DB 421 GCGAATGATCGGAGCGGCAATGCAATGCAACCTTAATCTCGCGCTGATCGCTGCGC 480
QY 780 GGTGTCTGCGCCACCTTCCCGGATGACGCAAACTTAAGAGAGAGCGCTGCGCTTC 839
DB 481 GGTGTCTGCGCCACCTTCCCGGATGACGCAAACTTAAGAGAGAGCGCTGCGCTTC 540
QY 840 GCGCAACATCAATTTGGGCACTGACTGAGCGCTGCGATAGCGAGCAGCAGCTTCTAC 899
DB 541 GCGCAACATCAATTTGGGCACTGACTGAGCGCTGCGATAGCGAGCAGCAGCTTCTAC 600
QY 900 GCGCCGAGAGCCCTGTGTGCTGTGGGAGCTGGGAGAGCCCGGAGAGGCGCTTCAAGCGG 959
DB 601 GCGCCGAGAGCCCTGTGTGCTGTGGGAGCTGGGAGAGCCCGGAGAGGCGCTTCAAGCGG 660
QY 960 GACTGGGGCTTATCTATCTCCCAAGTTCGCAAGCTGATAGCTAGGCGCCCAAGCTTCA 1019
DB 661 GACTGGGGCTTATCTATCTCCCAAGTTCGCAAGCTGATAGCTAGGCGCCCAAGCTTCA 720
QY 1020 TTGAGAGATTCCTGAGCTGAGAGTGCAGAGCTCCCATCTGTCTGCTCCGAGGAGC 1079
DB 721 TTGAGAGATTCCTGAGCTGAGAGTGCAGAGCTCCCATCTGTCTGCTCCGAGGAGC 780
QY 1080 CTGGTGTCTCAAGACTTCTTGTGA 1103
DB 781 CTGGTGTCTCAAGACTTCTTGTGA 804
```

RESULT 2  
US-08-910-973-12  
Sequence 12, Application US/08910973  
Patent No. 5795723

GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson KindnessP/LC

```
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910, 973  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239, 238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: PHCR-1-10958  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 20A1 (neuroD3)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 55..768  
US-08-910-973-12
```

Query Match 10.1%; Score 147.6; DB 1; Length 1268;  
Best Local Similarity 71.2%; Pred. No. 1.1e-29;  
Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
QY 647 GCGGCGGAGAGCGCCACAGGCCCAAGAGGAGTGTGCACTGAGAGCAGAGCGAG 706
DB 270 GCGGCGGAGAGCGCCACAGGCCGCTCGAGAGCGCTGCACTGAGAGCGAGAG 329
QY 707 CCGGCGCAAGAGCGCAAGCGAGCGAGCGCAACCGCATGAGCAACTTAACTCCGCGCT 766
DB 330 CCGGCGCGCTCAAGCGCAAGCATCGGAGCGCAACCGCATGAGCAACTTAACTCCGCGCT 389
QY 767 GAGTGCCTGCGCGGATGTCTGCGCCACTTCCCGATAGCGCAACTTAAAGATGCA 826
DB 390 GAGTGCCTGCGCGAGCGATGTCTGCGCTCGTCCCGAGCAGCACTACCAAAATGCA 449
QY 827 GACCTGTGCTTCCCGCCCAACTATTTGGGCACTGAGAGCGGATAGCGGAGCGGA 886
DB 450 GACGCTGCGCTTCCCGCTTCAACTATGTGGGCTGCGCGAGAGCACTGCGCTGCGGAG 509
QY 887 CCAGAGCTTCTACGCGCCCGAGCGCGCTGTGCGC 920
DB 510 TCAAGGCTGCCCGGAGCGGATGCCCGGAGCGC 543
```

RESULT 3  
US-09-499-227-12

Sequence 12, Application US/09499227  
Patent No. 644463  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/499,227  
FILING DATE: 05-August-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-May-1994  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-May-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,973  
FILING DATE: 07-August-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheinnes, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FPCR-1-12742  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 20A1 (neuroD3)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 55..768  
US-09-499-227-12

Query Match 10.1%; Score 147.6; DB 4; Length 1268;  
Best Local Similarity 71.2%; Pred. No. 1.1e-29;  
Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGGCGGAGGCGGCAAGGCGGCAAGGAGTTGACAGTGAAGCAAGCAAGCAAG 706  
DB 270 GCGGGCGGAGGCGGCAAGGCGGCGGTCCTCGAGGCGCTGCTGCACTGCGGAGG 329  
QY 707 CCGGCGAAGAGGCGCAAGCAAGGCGGAGCGCAACCGATGCAACAATTAACTCCGCGCT 766  
DB 330 CCGGCGGCTCAAGGCGCAAGATCGGAGGCAACCGATGCAACAATTGAAGCGGCGCT 389  
QY 767 GGATCGCTGCGGCTGCTGCGCAACTTCCCGGATGACGCAAACTTACAAAGATGCA 826  
DB 390 GGAGCGACTGCGCAAGCTGCTGCTCTCTTCCCGGAGCAACCAAGCTACCAAAATGCA 449

QY 827 GACCTGCGCTGCGCCCAACAATTGAGCACTGACTCAGACGCTGGCATAGGGA 886  
DB 450 GACGCTGCTTGGCTTCAACAATCTGAGGCTGCGGAGCACTGGCGGCA 509  
QY 887 CCACAGCTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 920  
DB 510 TCAAGGCTGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 543

RESULT 4  
US-08-722-570-12  
Sequence 12, Application US/08722570  
Patent No. 655337  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1527 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-12

Query Match 9.9%; Score 145.2; DB 4; Length 1527;  
Best Local Similarity 69.2%; Pred. No. 4.8e-29;  
Matches 198; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 635 GAAGCTCGTGGGCGGCGGAGGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGG 694  
DB 456 GAGAGGCGGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 515  
QY 695 GACGCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 754  
DB 516 GCTGCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 575  
QY 755 TAACTCGGCGTGAATGCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 814  
DB 576 CAAAGCTGCGTGAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 635  
QY 815 TACAAGATGAGACCGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 874  
DB 636 CACCAAGATTAGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 695  
QY 875 GCGCATAGCGGACCAAGCTTCTACGCGCGGCGGCGGCGGCGGCGGCGGCGG 920





Oy 815 TACAAAGATGAAACCCCTGGCTGCGCCACAACTACATTGGGACTGATCAGACGT 874  
 Db 487 CACCAAGATGAAACCGCTGGCTGCCCTACAACTACATCTGGGCGCCCTGGGTGAAACT 546  
 Oy 875 GCGCATAGCGGACCAACAGCTTTCTACGGCCCCGAGCCCTGTGCCCC 920  
 Db 547 GCGCTGGACAGATCAAGGGCTCCCGGGGGGACAGTGGCCGGAGGCC 592

## RESULT 9

```

US-09-499-227-21
: Sequence 21, Application US/09499227
: Patent No. 6444463
:
: GENERAL INFORMATION:
: APPLICANT: Tapscott, Stephen J.
: APPLICANT: Olson, James M.
: TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
: STREET: 1420 Fifth Avenue, Suite 2800
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98101-2347
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/499,227
: FILING DATE: 05-August-1998
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/239,238
: FILING DATE: 06-May-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US95/05741
: FILING DATE: 08-May-1995
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/17532
: FILING DATE: 30-October-1996
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/910,973
: FILING DATE: 07-August-1997
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Sheiness, Diana K.
: REGISTRATION NUMBER: 35,356
: REFERENCE/DOCKET NUMBER: FPCR-1-12742
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 206-682-8100; 206-224-0735 (direct)
: TELEFAX: 206-225-0779
:
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1333 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Mus musculus
: IMMEDIATE SOURCE:
: CLONE: neuroD3
: FEATURE:
:
: NAME/KEY: CDS
: LOCATION: 101..835
:
: US-09-499-227-21

```

Query Match	9.6%	Score 140.4	DB 4	Length 1333
Best Local Similarity	68.2%	Pred. No. 8.6e-28		
Matches 195, Conservative	0	Mismatches 91	Indels 0	Gaps 0

635 GAAGCTCCCTGGCGCGCGCGAGAGGGCGCAACAGCCCAAGAGCGAATTGGCACTAGACCA 694

Db	307	GCAGGAACGGCGGAGCGCGAGGCTCGGCTCGGGTGGCGGCTCCGAGGCTCTGCTGCACCTC	366
Oy	695	GCAGCGACGAACCGCGCGCAGAAAGGCAAGACCGGGAGCGCAACCGCATGCACAACT	754
Db	367	CTGCGGAGAGAGTCTGTGCTCAAGCCAAACGATGGAGCGCAACCGCATGCACAACT	426
Oy	755	TAACTCCGCTGGATGCGCTGCGCGGTGTCCTGCCACTTCCCGATGAGCGCAACT	814
Db	427	CAACGTGGCGCTGGACGCGCTTCGCGAGGCTCTGCGCTCTGTTCCCGACGACAAAGCT	486
Oy	815	TACAAAGATCGAGACCTCGCTTCGCTGCCACAACTACTTTGGGCACTGACTCAGAGCT	874
Db	487	CACCAAGATTGAGAGCGCTGCGCTTGCGCTAACACTGACTTGCGCCCTGGCTGAGACCT	546
Oy	875	GGCGATAGCGGACACACGTTTAAAGGCCCGGAGCGCCCTCTGGCC	920
Db	547	GGCGCTGGACAGATCAAGGGCTCCCGGGGGGACAGTCCCGGAGAGGC	592

## RESULT 10

US-08-932-411A-17  
Sequence 17, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Flahir Hohbach Test Albritton & Herdert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:

```

; LOCATION: 382..1170
US-08-932-411A-17

Query Match          9.6%; Score 139.8; DB 4; Length 1385
Best Local Similarity 78.1%; Pred. No. 1.3e-27;

```

Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 705 AGCCGCGCAGAAAGGCAAGCAGCGGAGGCAACCGCATGCAACCTTAACCTCCGCG 764  
Db 715 ACCCGCAGCTCAAGGCGCAACCGGAGGCAACCGCATGCAACCTTAACCGCGCG 774  
Qy 765 CTGATGCGCTGCGCGGCTGCTCTGCGCACTTCCCGATGAGCGCCAACTTAAGATC 824  
Db 775 CTGACGCGCTGCGGAGGTGCTGCGCACTTCCCGAGATGCGCACTGACGAGATC 834  
Qy 825 GAGACCTGCGCTTCCCGCAACATCACTTTGGGACATGATCAAGCTGCGCATAGCG 884  
Db 835 GAGACGCTGCGCTTCCCGCAACATCACTTTGGGACATGATCAAGCTGCGCATAGCG 894  
Qy 885 GACCACAGCTTCTAGCGCGCGGAGCGCGCTGCGC 919  
Db 895 GACCACTGCGCGCGCGCGCTGCGCTTCCAGGCGCG 929

RESULT 11  
US-08-552-142A-12  
Sequence 12, Application US/08552142A  
Patent No. 5695995  
GENERAL INFORMATION:  
APPLICANT: Weintraub, Harold M.  
APPLICANT: Lee, Jacqueline E.  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Hollenberg, Stanley M.  
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/552.142A  
FILING DATE: 02-NOV-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FPCR-1-8933  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-225-0709  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 20A1  
FEATURE:  
NAME/KEY: CDS

LOCATION: 1..310  
US-08-552-142A-12  
Query Match 8.3%; Score 121.8; DB 1; Length 310;  
Best Local Similarity 65.2%; Pred. No. 4.6e-23;  
Matches 191; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

Qy 619 GCCGAGGAGACATCGAGAGCTCCGTGCGCGCGGAGGCGGCAAGCGCCAGAGCG 678  
Db 16 GTCGAGGGGACACAGAGAGAGAGAGAGGCGCGCGCGCGGAGCGCGAGTCCCTCC 75  
Qy 679 AGTTGGCATGAGGACGAGGAGGAGGAGGCGCGCGGAGGAGGAGGAGGAGGAG 738  
Db 76 AGCGCTGCTGCAACCGCTGCGCGAGAG-CGGCGCTGCAAGGAGGAGGAGGAGGAG 134  
Qy 739 ACCGATGACCAACCTTAACCTCGCGCTGAGATGCGGCGGAGTGTCTGCGCACTTC 798  
Db 135 ACCGATGACCAACCTTAACCTCGCGCGCTGAGAGGAGGAGGAGGAGGAGGAGGAG 194  
Qy 799 CGGATGACGCGCAACTTAACAAGATCGAGACCTGCGCTTCCGCGCAACTATTGGG 858  
Db 195 CCGAGACACCAACTCACCAGAAATCGAGAGCTGCGCTTCCGCTTACCAACTATCTGG 254  
Qy 859 CACTGACTGAGCGCTGCGCATAGCGGAGCACAGCTTCTACGCGCCGAGGCC 911  
Db 255 CTCTGCGCGAGACACTGCGCTGCGCGATMAAGGCGCTGCCGAGGCGGTGCC 307

RESULT 12  
US-08-722-570-14  
Sequence 14, Application US/08722570  
Patent No. 655337  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722.570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-14

Query Match 7.9%; Score 115.6; DB 4; Length 1312;  
Best Local Similarity 67.4%; Pred. No. 3.1e-21;  
Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGCGAGGCGGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACGACGACGAG 706  
DB 539 GAGAGCGCGAGCGCGCGCTCAGGCGCAAGCGGAGAACTGTGTTAAAGATCAAGAAAC 598  
QY 707 CCGGCGCAAGAGCGCAACGAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCGCT 766  
DB 599 CCGGCGCGTTAAAGCTAAACAACCGGGAAGAAATCGCATGCAACCTGAACCTCGCGCT 658  
QY 767 GGATGCGGTGGCGGCTGTCTCCCTCCACCTTCCCGGATGACGCAAACTTACAAAGATGCA 826  
DB 659 TGATTCCTCAGGAGAGGTGTGCTCTTACCTGAAGATGCAAACTCACCAGATGCA 718  
QY 827 GACCTGCGCTTCGCCCAACTACATTGGGCACTGACGCTGCGCATAGCGGA 886  
DB 719 GACCTGCGCTTTCGCTTCAACTACATCTGGGCTCTTACGCAAACTTTGGCGCTGGGGA 778  
QY 887 CC 888  
DB 779 CC 780

RESULT 13  
US-08-932-411A-15  
Sequence 15, Application US/08932411A  
Patent No. 6566496

## GENERAL INFORMATION:

APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qinfu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr Hobach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1312 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-932-411A-15

Query Match 7.9%; Score 115.6; DB 4; Length 1312;  
Best Local Similarity 67.4%; Pred. No. 3.1e-21;  
Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGCGAGGCGGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACGACGACGAG 706  
DB 539 GAGAGCGCGAGCGCGCGCTCAGGCGCAAGCGGAGAACTGTGTTAAAGATCAAGAAAC 598  
QY 707 CCGGCGCAAGAGCGCAACGAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCGCT 766  
DB 599 CCGGCGCGTTAAAGCTAAACAACCGGGAAGAAATCGCATGCAACCTGAACCTCGCGCT 658  
QY 767 GGATGCGGTGGCGGCTGTCTCCCTCCACCTTCCCGGATGACGCAAACTTACAAAGATGCA 826  
DB 659 TGATTCCTCAGGAGAGGTGTGCTCTTACCTGAAGATGCAAACTCACCAGATGCA 718  
QY 827 GACCTGCGCTTCGCCCAACTACATTGGGCACTGACGCTGCGCATAGCGGA 886  
DB 719 GACCTGCGCTTTCGCTTCAACTACATCTGGGCTCTTACGCAAACTTTGGCGCTGGGGA 778  
QY 887 CC 888  
DB 779 CC 780

RESULT 14  
US-08-722-570-15  
Sequence 15, Application US/08722570  
Patent No. 655337

## GENERAL INFORMATION:

APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qinfu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESS: Flehr, Hobach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-15

Query Match 7.5%; Score 109.8; DB 4; Length 1277;  
Best Local Similarity 74.6%; Pred. No. 1e-19;  
Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCGGCGCGAAGAGCGCAACGCGGAGCGCAACCGCATGCAACCTTAACCTCGCG 764  
DB 439 ACCGCGCGCGTTAAAGCGCAAAATACCGGAGAGATCGCATGCAACCTTAACCTCGCG 498  
QY 765 CTGATGCGCTGCGGCGGTCTCTCCGCCACCTTCCCGATGACGCCAACTTACAAAGATC 824



Db 499 CTCGATTCTGAGGAGGTTCTACCGTCATTACCCGAGCGCAACCTCACCAATA 558  
QY 825 GAGACCTGCGCTTGGCCACACATCACTTTGGGCACTGACTGAGCGCTGGCATAGCG 884  
Db 559 GAGACCTTGGCGCTTGGCCACACATCACTTGGGCTCTTACCGAACTTTGGCGCTGCGC 618  
QY 885 GACCA 889  
Db 619 GACCA 623

Db 559 GAGACCTTGGCGCTTGGCCACACATCACTTGGGCTCTTACCGAACTTTGGCGCTGCGC 618  
QY 885 GACCA 889  
Db 619 GACCA 623  
Search completed: January 26, 2004, 22:06:42  
Job time : 100 secs

RESULT 15  
US-08-932-411A-16  
; Sequence 16, Application US/08932411A  
; Patent No. 6566496  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David J.  
; APPLICANT: Ma, Qifu  
; TITLE OF INVENTION: NEUROGENIN  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/932,411A  
; FILING DATE: 15-SEP-1997  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/772,009  
; FILING DATE: 19-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/722,570  
; FILING DATE: 19-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-63902-3/RPT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1277 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; US-08-932-411A-16

Query Match 7.5%; Score 109.8; DB 4; Length 1277;  
Best Local Similarity 74.6%; Pred. No. 1e-19;  
Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
QY 705 AGCCGGCGCAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGG 764  
Db 439 ACCCGGGCGGCTTAAGCAATACCGGAGAGGAGATCGCATGACACCTGACTATGCG 498  
QY 765 CTGATGCGCTGCGCGGCTGCTCTGCGGCACTTTCCGGAGTACGCGCAACTTACAAAGTC 824  
Db 499 CTGATGCTTGAGGAGGAGTCTACCTGCTTACCGGAAAGCCCAACTCACCAAGATA 558  
QY 825 GAGACCTGCGCTTGGCCACACATCACTTTGGGCACTGACTGACAGCGCTGGCATAGCG 884

**THIS PAGE BLANK (USPTO)**



Db 714 ACTGCATGACGCTGAGAAATCCCTGTGGTCTCATCACTGACGAGCTGTGAGTACT 773  
Qy 178 CCTCGAGCTCTTGTGCTGCTCCGAGAGCAATTACTCCAGGGGCGCTGACCT 237  
Db 774 CCTGGAGCTTTTACGACTTCCAGACGCAATTACTCCAGGGGCGCTGACCT 833  
Qy 238 CAGCAAACTTCCAGAGCGAGAGGGGTTCACTATCCACCGCTGTGACTGACC- 296  
Db 834 TAGCAAACTTCCAGAGGGGAGAGAGGCTATCCACTGCTGTGACACTGACC 893  
Qy 297 -----ACCCGAGCTCTCTGTTCTTTTGAAGCCGAGTA 330  
Db 894 TATCCACTGCTGCTTGTCACTGACTGCTGCTCTTCTTATCTTTGAGTGGAG-A 952  
Qy 331 ACTAGTAACATTAGGAACCTCCAAAGGGTGAAGAGGGGAGTGGGT-----GG 380  
Db 953 ACTAGTAACATTAGGAACCTCCAAAGGGTGAAGAGGGGAGGCGGGGCTGTGGG 1012  
Qy 381 GCGTACTGATCCCGCTGAGTACCTTAAGTCAAGACTG--TCAACCCCTTC 438  
Db 1013 GGATCTCTGCTCCCGCTGAGTACCTTAAGTCAAGAGGCTGGCAGACACACTTC 1072  
Qy 439 CATTTTTCCTCAACTCAGATGCGGCTCATCTCTTGATGCGCCCACTCAAGTGT 498  
Db 1073 CATTTTTCCTCAACCGAGATGCGGCTCATCTCTTGATGCGCTCACTCAAGTGT 1132  
Qy 499 CCCAGAGACCCAGAACCTTTCCCGAGCCTCGAGACCAAGATGCTCACTTCCATT 558  
Db 1133 CCCAGAGACCAACCTTTTCCCGAGCCTCGAGACCAAGATGCTCACTTCCATT 1192  
Qy 559 CCACCCCACTTACCTCTCTGTAACGAGGGACTGCTCCGAGCAGAGGAGTACT 618  
Db 1193 CCACCCCACTTACCTCTCTGTAACGAGGGACTGCTCCGAGCAGAGGAGTACT 1252  
Qy 619 GCCGAGGACATCGAGAACTCTGCTGCGGCGCGAGGAGGCGCAACAGGCCAAGAG 678  
Db 1253 GCCGAGGAGCTCGAGAACTCTGCGGCGCGAGGAGGCGCAACAGGCCAAGAG 1312  
Qy 679 AGTTGGCACTGAGCAAGCAGAGGCGGAGGCGCAAGAGGCGCAAGCGGGAGCGCA 738  
Db 1313 AGTTGGCACTGAGCAAGCAGAGGCGGAGGCGCAAGAGGCGCAAGCGGGAGCGCA 1372  
Qy 739 ACCGATGACAACTTAACTCCGCTGAGTGTGCTGCGGCTGCTGCTGCTGCTGCT 798  
Db 1373 ATCGATGACAACTTAACTCCGCTGAGTGTGCTGCGGCTGCTGCTGCTGCTGCT 1432  
Qy 799 CGGATGACGCAACTTACAAAGATGAGACCTTGTGCTGCGCCCACTCAATTGGG 858  
Db 1433 CGGATGACGCAACTTACAAAGATGAGACCTTGTGCTGCGCCCACTCAATTGGG 1492  
Qy 859 CACTGATCAGACGCTGCGCATAGGAGACCAAGCTTCAAGCCCGAGCCCTGCTGTC 918  
Db 1493 CACTGATCAGACGCTGCGCATAGGAGACCAAGCTTCAAGCCCGAGCCCTGCTGTC 1552  
Qy 919 CTTGTGGGAGTGGGAGCCCGGAGGAGGCTCCAGCGGCACTTGGGCTCTATCTACT 978  
Db 1553 CTTGTGGGAGTGGGAGCCCGGAGGAGGCTCCAGCGGCACTTGGGCTCTATCTACT 1612  
Qy 979 CCCCACTTCCCAAGCTGTAGCTGAGCCCAAGCTTCAATTGAGAGATTCCCTGTC 1038  
Db 1613 CCCCACTTCCCAAGCTGTAGCTGAGCCCAAGCTTCAATTGAGAGATTCCCTGTC 1672  
Qy 1039 TCGAGTGGCCAGCTCTCCATCTGCTGCTCCGAGGACCTGCTGCTTCAAGCTTCT 1098  
Db 1673 TCGAGTGGCCAGCTCTCCATCTGCTGCTCCGAGGACCTGCTGCTTCAAGCTTCT 1732  
Qy 1099 TCGAGAGGCGCAAGAGCCCTGCGGAGGAGGCTGCGCAAGAGAGGAGGAGTCAAG 1158  
Db 1733 TCGAGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1792  
Qy 1159 CTGTCTGAATGAGAGTGAAGGAGCACTGAGCACTTCTGCTGCTTCAATT 1218  
Db 1793 CTGTCTGAATGAGAGTGAAGGAGCACTGAGCACTTCTGCTGCTTCAATT 1852

Qy 1219 GTGAGTCC 1227  
Db 1853 CTMGATTC 1861

## RESULT 2

US-10-004-717-24  
; Sequence 24, Application US/10004717  
; Publication No. US2002019265A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGBIL, HUDA Y.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATOUAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P01899054  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; PRIOR FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 861  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-717-24

Query Match 46.7%; Score 682.4; DB 14; Length 861;

Best Local Similarity 89.7%; Pred. No. 1.5e-185; Indels 13; Gaps 3;  
Matches 771; Conservative 0; Mismatches 76;

Qy 312 TTCTTTGAGCCCGAGTAACATTAGTAACCTTCAAGAGGTAAGAGGG 371  
Db 2 TTCTTTGAGTGGAGAG-AACTAGTAACATTGGAAACTCCAAAGGTTGATGAGGG 60  
Qy 372 AGTGGT-----GGGCGTACTGATGCTGCGGCTGAGTGACTTAAAGTCAAGA 421  
Db 61 CGCGGAGGTGTGTGTGGGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Qy 422 CTG--TCAACCCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 479  
Db 121 CTGGCAACACACACCTTCCATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 180  
Qy 480 GCGCCACATCCAAAGTGTCCCAAGAGACCGAGCAACCTTCCCGAGGCTCGAGCCAC 539  
Db 181 GCGCTACATCCAAAGTGTGTCCCAAGAGACCGAGCAACCTTCCCGAGGCTCGAGCCAC 240  
Qy 540 GAAGTGTCTCAATTCATTCACCCCACTGACCTTCTGTAACGAGGAGTGTCTTC 599  
Db 241 GAAGTGTCTCAATTCATTCACCCCACTGACCTTCTGTAACGAGGAGTGTCTTC 300  
Qy 600 GAAGCAAGAGAGTGTGCTGCGAGGAGCATGAGAGAGTCTGCTGCGGCGCGGAGGG 659  
Db 301 GAAGCAAGAGAGTGTGCTGCGAGGAGCATGAGAGAGTCTGCTGCGGCGCGGAGGG 360  
Qy 660 GCGCAAGGCGCAAGAGGAGTGTGCACTGAGCAAGAGCGAGGAGCGGAGGAGGAG 719  
Db 361 GCGCAAGGCGCAAGAGGAGTGTGCACTGAGCAAGAGCGAGGAGCGGAGGAGGAG 420  
Qy 720 GCGCAAGGCGGAGGAGCGCAAGGAGTGTGCACTGAGCAAGGAGGAGGAGGAGGAG 779  
Db 421 GCGCAAGGCGGAGGAGCGCAAGGAGTGTGCACTGAGCAAGGAGGAGGAGGAGGAG 480  
Qy 780 GGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 839  
Db 481 GGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540



```

Db      2775 CTGAACCTTGGGACGACGAAAGCCCGCTGAGCTCCCAACGGCCCTCGCTGATCGCTCTC 2834
Qy      286 TGACTCTGACCAACCCGACAGCTCTCTGTTCTTTGAGCCCGAGTAAGTAACATTTA 345
Db      2835 TATCTTTTGGCCCGGTGAAAGGTAATTTTGGAGGCTCCGAGGAGCGGGCAGGGAA 2894
Qy      346 GGAAC-CTCCAAAGGTAAGAGGAGGAGTGGGTGGGCTACTAGTCCGCGTGAAGT 404
Db      2895 AGAGGAGATCTCTGACCCAGCGGGGCTGGAGAGATGGCTGTTTTTTTTTTTCCCACT 2954
Qy      405 GACCTTAAGTCAAGACTG-----TGACACCCCTTCATTTTTCGAA 451
Db      2955 AGCCTCGAATCGCGACTGCGCCGCTGACGAGCTCAAACTTACCTTCCTCTGACCCCG 3014
Qy      452 CCTGAGATGGGCTCTATCCCTTTGATGCGCCCAACATCCAAAGTCTCCCAAGAACCA 511
Db      3015 CCGTAGATGAGGCTTCAACCTTCGGGTGCGCCCACTGTCCAAAGTACCCGTGAGACGA 3074
Qy      512 GCAACCTTTCCCGAGCTCGGACCAAGAGTGTCAAGTTCCAAATCCACCCCACTAG 571
Db      3075 GCGGTCTTTCCTCCAGAGCTTCGGAAGAGAGAGTGAAGTCCCACTCCGCGCCCGCAG 3134
Qy      572 CCCCCTCTCTGTAACGAGGACTGTCTCCGAGCAGAGAGAGTGAAGTCCGAGGAGCAATC 631
Db      3135 CCCCCTCGACACAGGGGGGAACTGGCGCAGAGCGGAAAGAGAGGAGGCTGCCAGGGGCCCC 3194
Qy      632 GAGGAAGCTCCGTCGCGCGCGCGGAGGGGCGCACAGGCCCAAGAGCGAGTTGGCACTGAG 691
Db      3195 GAGGAAGCTCCGCGGACCGCGCGGGGAGCGCAGCGGCGCTTAAGACGAGTTGGCACTGAG 3254
Qy      692 CAAGAGGAGGAGAGCGCGCGGACAGAGGCGCAAGCAGCGGGAGCGCAACCGCATGACAA 751
Db      3255 CAAGAGGAGGAGAGTGGCGGAAAGAGGCGCAAGCAGCGCGGAGCGCATGAGTGCACAA 3314
Qy      752 CTTTAACCTCCGCGCTGATGCGCTGCGCGGTGTCTCTCCCACTTCGCGATGAGCGCA 811
Db      3315 CTTCAACTCTGACAGAGCGCTGCGCGGTGTCTCTCCCACTTCGCGATGAGCGCA 3374
Qy      812 ACTTCAAAAGATCGAGAGCTTCGCTGCGTCCCACTTAATTGGGCACTGACTGAGC 871
Db      3375 GCTTCAAAAGATCGAGAGCTTCGCTGCGTCCCACTTAATTGGGCACTGACTGAGC 3434
Qy      872 GCTGCGATGAGCGGACAGCTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931
Db      3435 GCTGCGATGAGCGGACAGCTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3494
Qy      932 GGAAGCGCGGAGGCGGCTCCAGCGCGCACTGGGGCTTATCTACTCCCACTTTCCCA 991
Db      3495 GCTGGGCAAGCCAGGCGGTTCCCGCGGGGACTGGGGGTCTCTACTCCCACTTTCCCA 3554
Qy      992 AGCTGTAGCTGAGAGCCCAAGCGCTCATTTGAGAGATTCCTGGCGCTGAGAGTCCCG 1051
Db      3555 GCGTGGCAGCTGAGAGTCCCGCGCGCTGAGAGGAGCAACCGGCGCTGCTGGGGCGCAC 3614
Qy      1052 CTCCCATCTGCTGCTGCTCCCGGCAACCTGAGTTCTGAGCTTCTGAAAGGCGCA 1111
Db      3615 CTCTTCGCTGCTGAGAGCTGAGGCTTGGCTTTCTAGATTTTCTGTAAGAGACTT 3674
Qy      1112 AACAGCGCTGGGCGGTGGCGCTGGCAGAAAGAGGAGAGTCAAGCTGTCTGAATGG 1171
Db      3675 GTCTGTGCTGGGCTGTGGGTCTTAAGGTAAAGGAGAGGAGAGCGGGAGCGGTAG 3734
Qy      1172 AAGGTAGTGAAGCACTGAGCATC 1196
Db      3735 AGGTGGCGGACGCGCGCGCTC 3759

```

RESULT 5  
 US-08-722-570-12  
 ; Sequence 12, Application US/08722570  
 ; Publication No. US20030044887A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
;
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1527 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
;
; US-08-722-570-12

```

Query Match 9.9%; Score 145.2; DB 8; Length 1527;  
 Best Local Similarity 69.2%; Pred No. 2e-31; Indels 0; Gaps 0;  
 Matches 198; Conservative 0; Mismatches 88;

```

Qy      635 GAAAGCTCCGTGCGCGCGCGGAGGCGGCAAGAGCGGAGGAGTGGCACTGAGCA 694
Db      456 GAAGAGGCGGCGGCGACGCGAGGTGCGCGGGGTGCGGTCCGAGGCGCTGCTCACTC 515
Qy      695 GCAGGAGAGAGCGGCGGCGGCAAGAGCGGAGCGGCAACCGCATGCAAACT 754
Db      516 GCTGGGAGGAGCGCTGCTGCAAGCGCAAGATGCGAGGCAACCGTATGCAATACCT 575
Qy      755 TAACTCCGCGCTGATGCGCTGCGCGGTGCTGCTGCGCACTTCGCGATGAGCGCAACT 814
Db      576 CAAGCTGCGCTGAGAGCGCTGCGAGAGGTGCTGCTGCTTCCCGAGACCAAGACT 635
Qy      815 TACAAAGATGAGAGCCCTGCGCTTGCCCACTAATTGGGCACTGACTCAGAGCT 874
Db      636 CACCAAGATTGAGAGCTGCGCTTGCTGCTTACAACTAATCTGAGGCGCTGCTGAGACACT 695
Qy      875 GCGCATAGCGGACCACTTCTAGCGGCCCGAGGCCCTGTGCCCC 920
Db      696 GCGCTGCGAGATCAAGGCTTCCCGGGGCGGTGCTCCGGAGGCGC 741

```

RESULT 6  
 US-08-722-570-13  
 ; Sequence 13, Application US/08722570  
 ; Publication No. US20030044887A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Anderson, David J.  
 ; APPLICANT: Ma, Qifu  
 ; TITLE OF INVENTION: NEUROGENIN  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/722,570  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-13

Query Match 9.6%; Score 140.4; DB 8; Length 738;  
Best Local Similarity 68.2%; Pred. No. 4,1e-30;  
Matches 195; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 635 GAAGCTCCGCGCGCGCGGAGGCGGCAAGAGCCCAAGAGCGATTGGACCTGAGCAA 694  
DB 207 GCAGAGAGCGGCGGAGCGCGGAGGTCGCGTGGGTGGCGTCCAGGCTGCTGCTCACTC 266  
QY 695 GCAGCGAGAGCGCGCGGCGGAGGCGGCAAGAGCGGAGCGGCAAGCGGATGCAACT 754  
DB 267 CCTGGGAGGAGTCTCGCTGCAAGCGGATGCGAGGCGCAAGCGGATGCAACT 326  
QY 755 TAACCGCGCGTGGATGCGTGGCGGCGGATGCTGCTGCGCACTTCCGAGTGAAGCCAACT 814  
DB 327 CAACGCTGCGTGGAGCGCTTGGCGAGCGTGTGCTGCTTCTTCCCGAGCGACCAAGCT 386  
QY 815 TACAAAGATCGAGAGCGCTTGGCGGCGGCAAGCGGATGCGGAGCGGATGCAACT 874  
DB 387 CACCAAGATGAGAGCGTGGCGGCTTGGCGGCGGATGCGGCGGCGGCGGCGGCGGCGG 446  
QY 875 GCGCATAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 920  
DB 447 GCGCGTGGCGAGTCAAGGCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 492

RESULT 7  
US-10-004-717-30  
Sequence 30, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHBI, HUDA Y.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899US4  
CURRENT FILING DATE: 2002-08-16  
CURRENT APPLICATION NUMBER: US/10/004,717  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 30  
LENGTH: 1385  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-30

Query Match 9.6%; Score 139.8; DB 14; Length 1385;  
Best Local Similarity 78.1%; Pred. No. 7e-30;  
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGGCGCAAGCGGAGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 764  
DB 715 ACCCGAGCGCTCAAGCGGCAAGCGGAGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 774  
QY 765 CTGATGCGCTGCGCGGCTGCTGCTGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 824  
DB 775 CTGAGCGCGCTGCGCGGAGGCTGCTGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 834  
QY 825 GAGACCTGCGCTTCCCGGCAAGCGGATGCAAGCGGATGCAAGCGGATGCAAGCGG 884  
DB 835 GAGACGCTGCGCTTCCCGGCAAGCGGATGCAAGCGGATGCAAGCGGATGCAAGCGG 894  
QY 885 GACCAGAGCTTCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 919  
DB 895 GACCACTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 929

RESULT 8  
US-10-004-717-6  
Sequence 6, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHBI, HUDA Y.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899US4  
CURRENT FILING DATE: 2002-08-16  
CURRENT APPLICATION NUMBER: US/10/004,717  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1412  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-6

Query Match 9.6%; Score 139.8; DB 14; Length 1412;  
Best Local Similarity 78.1%; Pred. No. 7e-30;  
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGGCGCAAGCGGAGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 764  
DB 418 ACCCGAGCGCTCAAGCGGCAAGCGGAGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 477  
QY 765 CTGATGCGCTGCGCGGCTGCTGCTGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 824  
DB 478 CTGAGCGCGCTGCGCGGAGGCTGCTGCGGCAAGCGGATGCAAGCGGATGCAAGCGG 537  
QY 825 GAGACCTGCGCTTCCCGGCAAGCGGATGCAAGCGGATGCAAGCGGATGCAAGCGG 884  
DB 538 GAGACGCTGCGCTTCCCGGCAAGCGGATGCAAGCGGATGCAAGCGGATGCAAGCGG 597

OY 885 GACCAAGCTTCTACGGCCCGAGCCCTGTGCC 919  
DB 598 GACCACTGGCGCGCGCGGTGGCTTCCAGGGGGC 632

## RESULT 9

US-10-004-717-37  
Sequence 37, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:

APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899US4  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37  
LENGTH: 1412  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-37

Query Match 9.6%; Score 139.8; DB 14; Length 1412;  
Best Local Similarity 78.1%; Pred. No. 7e-30;  
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY 705 AGCCGCGCAAGAGGCAAGCAAGCGGAGCGCAACCGCATGCACAACTTAATCCGGC 764  
DB 418 ACCCGAGGCTCAAGGCAACCAACCGGCAACCGCATGCACAACTTAATCCGGC 477  
OY 765 CTGATGCGCTGCGCGGTGCTCTGCCACCTTCCCGATGACGCCCAACTTAAGATC 824  
DB 478 CTGAGCGCGCTGCGGAGGTGCTGCCACCTTCCCGAGATGCCAAGCTCAAGATC 537  
OY 825 GAGACCTGCGCTGCGGCAAGCAATGATGCGCATGCTGAGCGGTGGCATAGG 884  
DB 538 GAGAGCGCTGCGCTGCGGCAATGATGCTGAGCGGTGGCATAGG 597  
OY 885 GACCAAGCTTCTACGGCCCGAGCCCTGTGCC 919  
DB 598 GACCACTGGCGCGCGCGGTGGCTTCCAGGGGGC 632

## RESULT 10

US-10-004-717-20  
Sequence 20, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:

APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899US4  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01

NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 790  
TYPE: DNA  
ORGANISM: chicken  
US-10-004-717-20

Query Match 9.5%; Score 139.2; DB 14; Length 790;  
Best Local Similarity 84.8%; Pred. No. 9.3e-30;  
Matches 156; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 702 CGAAGCCGCGCAAGAGGCAAGCAAGCGGAGCGCAACCGCATGCACAACTTAATCC 761  
DB 344 CGAGCCGCGCGCTGAAGAGCAACCGGCAACCGCATGCACAACTTAATCC 403  
OY 762 GCGCTGATGCGCTGCGCGGTGCTCTGCCACCTTCCCGATGACGCCCAACTTAAG 821  
DB 404 GCGCTGAGCGCGCTGCGGAGGTGCTCTGCCACCTTCCCGAGAGCGCAAGCTCACCAAG 463  
OY 822 ATCGAGACCTGCGCTTGGCCCAACATTAATTTGGCATGACTGACAGCGCTGGCAT 881  
DB 464 ATCGAGAGCGCTGCGCTTGGCCCAACATTAATTTGGCGGCTCACCGAGACGCTGGCCTG 523  
OY 882 GCGG 885  
DB 524 GCGG 527

## RESULT 11

US-10-004-717-18  
Sequence 18, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:

APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899US4  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 18  
LENGTH: 1074  
TYPE: DNA  
ORGANISM: chicken  
US-10-004-717-18

Query Match 9.5%; Score 138.2; DB 14; Length 1074;  
Best Local Similarity 71.4%; Pred. No. 1.9e-29;  
Matches 182; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

OY 642 CGTGGCGCGCGCGGAGGCGCAAGCGCCCAAGAGCGATGGCACTGAGCAGCAGCA 701  
DB 372 CGGCGAGAGAGAGCGGCGCGGTGCGGCGGCGGAGCGGAGCTTTGCTGACACCTTCAA 431  
OY 702 CGAAGCCGCGCAAGAGGCAAGCAAGCGGAGCGCAACCGCATGCACAACTTAATCC 761  
DB 432 CGAGCCGCGCGGTGAAGCAAGCAAGCGGAGCGGAGCGGAGCGGATGACCACTCAAGCGC 491  
OY 762 GCGCTGATGCGCTGCGCGGTGCTCTGCCACCTTCCCGATGACGCCCAACTTAAG 821  
DB 492 GCGCTGATGAGCTCGAGAGCTTCTGCGGAGCTTCCCGAGAGCAGACCAACTTCAAG 551  
OY 822 ATCGAGACCTGCGCTTGGCCCAACATTAATTTGGGCACTGACTGACAGCGCTGGCAT 881



Db 552 ATCGAAACCTGCGCTTACCAATCTGGGCGCTCTCGAGACCTTCGTTTG 611  
QY 882 GCGAGCACAGCTTC 896  
Db 612 GCCAGCAGTGCCTC 626

## RESULT 12

US-08-722-570-14  
; Sequence 14, Application US/08722570  
; Publication No. US20030044887A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David J.  
; APPLICANT: Ma, Qifu  
; TITLE OF INVENTION: NEUROGENIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/722,570  
; FILING DATE: 27-SEP-1996  
; CLASSIFICATION: 5365  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1312 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; US-08-722-570-14

Query Match 7.9%; Score 115.6; DB 8; Length 1312;

Best Local Similarity 67.4%; Pred. No. 6e-23; Indels 0; Gaps 0;

Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGGAGGCGGCAAGAGGCGAGTGGCACTGACAGCAGCAGCAAG 706  
Db 539 GAGGAGCCGAGGCGGCTCAGGCGCAAGGCGGAGAACTGTAAAGATCAAGAAAC 598  
QY 707 CCGGCGAAGAGGCGCAAGCAGCGGAGCGCAACCGCATGCAACCTTAATCGCGCT 766  
Db 599 CCGGCGGCTTAAAGCTAACACCGGGAAGATGCGATGCAACCTGAATCTGCGCT 658  
QY 767 GGATGCGGTGGCGGTCTCTGCGCACTTCCCGATGACGCGCAACTTACAAAGATCGA 826  
Db 659 TGATTCCTCAAGGAGGTGTGCTCTTAACTTAAGATGCAAACTCACCAATAGA 718  
QY 827 GACCTGCGCTTGGCCCAACTATATTGGGCACTGACTGACGCTGGCATAGCGGA 886  
Db 719 GACCTTGGCTTGGCTTACACTATCTGGGCTCTTACGCAAACTTTGGCGCTTGGCGA 778  
QY 887 CC 888  
Db 779 CC 780

## RESULT 13

US-08-722-570-15  
; Sequence 15, Application US/08722570  
; Publication No. US20030044887A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, David J.  
; APPLICANT: Ma, Qifu  
; TITLE OF INVENTION: NEUROGENIN  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/722,570  
; FILING DATE: 27-SEP-1996  
; CLASSIFICATION: 5365  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Silva, Robin M.  
; REGISTRATION NUMBER: 38,304  
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 781-1989  
; TELEFAX: (415) 398-3249  
; TELEX: 910 277299  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1277 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA  
; US-08-722-570-15

Query Match 7.5%; Score 109.8; DB 8; Length 1277;

Best Local Similarity 74.6%; Pred. No. 2.8e-21; Indels 0; Gaps 0;

Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGGCGCAAGCAGCGGAGCGCAACCGCATGCAACCTTAATCTCGCG 764  
Db 439 ACCCGGCGCTTAAAGCCAAATACCGGAGAGATGCGATGCAACCTGAATATGCG 498  
QY 765 CTGATGCGCTGCGCGGTCTCTGCGCACTTCCCGATGAGCGCAACTTAAAGATC 824  
Db 499 CTCATCTCTGAGGAGGTCTACGCTCATTAACCGAAGCGCAAACTCACCAAGATA 558  
QY 825 GAGACCTTGGCTTGGCCCAACTATATTGGGCACTGACTGACGCTGGCATGCGG 884  
Db 559 GAGACTTGGCTTGGCCCAACTATATTGGGCTCTTACGCAAACTTTGGCGCTGCGC 618  
QY 885 GACCA 889  
Db 619 GACCA 623

## RESULT 14

US-10-413-358-27  
; Sequence 27, Application US/10413358  
; Publication No. US20030219894A1  
; GENERAL INFORMATION:  
; APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.  
; TITLE OF INVENTION: Induction of insulin-producing cells  
; FILE REFERENCE: GP55

Job time : 520 secs

;; CURRENT APPLICATION NUMBER: US/10/413,358  
;; CURRENT FILING DATE: 2003-04-15  
;; PRIOR APPLICATION NUMBER: JP2002-115064  
;; PRIOR FILING DATE: 2002-04-17  
;; NUMBER OF SEQ ID NOS: 27  
;; SEQ ID NO: 27  
;; LENGTH: 1099  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-413-358-27

Query Match 6.1%; Score 89; DB 13; Length 1099;  
Best Local Similarity 61.4%; Pred. No. 2.5e-15;  
Matches 143; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 650 GCGCGAGAGCGCCACAGCGCCCAAGAGCGAGTGGCACTGAGCAGCAGCGAGCGCG 709  
DB 258 GAGACGCGCGCCCAAGAGAGAGATGACTAAGGCTGCGTGAAGCGTTTAAATTGAG 317  
QY 710 GCGCAAGAGAGCGCAAGCAGCGGAGCGCAACCGCATGCAACCTTAATCCGCGCTGGA 769  
DB 318 AGCATGAAGGCTAACGCGCGGAGCGGAGCGCATGCAAGCGAGCTGAAGCGCGCTAGA 377  
QY 770 TGGCGTGGCGGCTGTCTTGGCCCACTTCCCGGATGACGCCAACTTAACAAGATCGAGAC 829  
DB 378 CAACCTGGCGCAGAGTGTGCTTGTCTATTTAAGACGAGAGCTGTCCAAATCGAGAC 437  
QY 830 CTTGGCGTTCCGCCACAACTATATTGGGCACTGACTGAGCGCTGGCGATNG 882  
DB 438 TCTGGCGTTGGCAAGAACTATCATCTGGGCTGTGCGAGATCTGGCGCTCAG 490

## RESULT 15

US-10-413-358-26  
;; Sequence 26, Application US/10413358  
;; Publication No. US20030219894A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.  
;; TITLE OF INVENTION: Induction of insulin-producing cells  
;; FILE REFERENCE: GPs5  
;; CURRENT APPLICATION NUMBER: US/10/413,358  
;; CURRENT FILING DATE: 2003-04-15  
;; PRIOR APPLICATION NUMBER: JP2002-115064  
;; PRIOR FILING DATE: 2002-04-17  
;; NUMBER OF SEQ ID NOS: 27  
;; SEQ ID NO: 26  
;; LENGTH: 1211  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-10-413-358-26

Query Match 6.1%; Score 89; DB 13; Length 1211;  
Best Local Similarity 61.4%; Pred. No. 2.5e-15;  
Matches 143; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 650 GCGCGAGAGCGCCACAGCGCCCAAGAGCGAGTGGCACTGAGCAGCAGCGAGCGCG 709  
DB 348 GAGACGCGCGCCCAAGAGAGAGATGACTAAGGCTGCGTGAAGCGTTTAAATTGAG 407  
QY 710 GCGCAAGAGAGCGCAAGCAGCGGAGCGCAACCGCATGCAACCTTAATCCGCGCTGGA 769  
DB 408 AGCATGAAGGCTAACGCGCGGAGCGGAGCGCATGCAAGCGAGCTGAAGCGCGCTAGA 467  
QY 770 TGGCGTGGCGGCTGTCTTGGCCCACTTCCCGGATGACGCCAACTTAACAAGATCGAGAC 829  
DB 468 CAACCTGGCGCAGAGTGTGCTTGTCTATTTAAGACGAGAGCTGTCCAAATCGAGAC 527  
QY 830 CTTGGCGTTCCGCCACAACTATATTGGGCACTGACTGAGCGCTGGCGATNG 882  
DB 528 TCTGGCGTTGGCAAGAACTATCATCTGGGCTGTGCGAGATCTGGCGCTCAG 580

Search completed: January 26, 2004, 22:15:27









ECORI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI Methylase. Site selected DNA was cloned into the pBAC3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 88 a 159 c 213 g 133 t  
ORIGIN

Query Match 32.4%; Score 473.4; DB 28; Length 593;  
Best Local Similarity 91.5%; Pred. No. 1.8e-105;  
Matches 535; Conservative 0; Mismatches 46; Indels 4; Gaps 3;

426 CACACCCCTTCATTTTTCACCACTCAGATGCGGCTCATCCCTGGATCGGCC 485  
586 CACACACCTTCATTTTTCACCACTCAGATGCGGCTCATCCCTGGATCGGCC 527  
486 ACCATCCAGTGTCCCAAGACCAACCTTTCCCGAGCTCGAGACCAAGATG 545  
526 ACCATCCAGTGTCCCAAGACCAACCTTTCCCGAGCTCGAGACCAAGATG 467  
546 CTCAGTTCATTCACCCCACTCAGATGCGGCTCATCCCTGGATCGGCC 605  
466 CTCAGTTCATTCACCCCACTCAGATGCGGCTCATCCCTGGATCGGCC 407  
606 GAAGCAGGTGATCTGCGGAGGACATGAGAGTCCGTCGCGCGGAGGCGCAAC 665  
406 GAAGCAGGTGATCTGCGGAGGACATGAGAGTCCGTCGCGCGGAGGCGCAAC 347  
666 AGGCCCAAGAGGAGTGTGCACTGAGCAAGCAGACAGGCGGCGCAAGAGCC 725  
346 AGGCCCAAGAGGAGTGTGCACTGAGCAAGCAGACAGGCGGCGCAAGAGCC 287  
726 GACCGGAGCGCAACCGATGCAACCTTAACTCCGCTGAGTGGCTCGCGGTC 785  
286 GATCGGAGCGCAACCGATGCAACCTTAACTCCGCTGAGTGGCTCGCGGTC 227  
786 CTGCCCACCTTCGCGAGTACGCGCAAACTTACAAAGATGAGACCTCGCTCCGC 845  
226 CTGCCCACCTTCGCGAGTACGCGCAAACTTACAAAGATGAGACCTCGCTCCGC 167  
846 AACTCATTTGGGCACTGATGAGCGCTGCGCATAGCGGACCAAGCTTACGGCC 905  
166 AACTCATTTGGGCACTGATGAGCGCTGCGCATAGCGGACCAAGCTTACGGCC 107  
906 GAGCCCTGTCCTGTCGCGGAGGCTGAGAGCGCGGAGGCGGCTTCCAGGCGCA 965  
106 GAGCCCTGTCCTGTCGCGGAGGCTGAGAGCGCGGAGGCGGCTTCCAGGCGCA 49  
966 GAGCTATCTACTCTCCCAAGTTTCCCAAGCTGCTGAGCGCC 1010  
48 GAGCTC-ATCACTCCCAAGTCTCCCAAGCGGGAACCTTACCGCCAC 6

RESULT 4  
CA945402 687 bp mRNA linear EST 30-DEC-2002  
LOCUS  
DEFINITION UT-M-FD0-cdh-1-12-0-UT.r1 NIH\_BMAP\_F00 Mus musculus cDNA clone  
IMAGE: 6828925 5', mRNA sequence.  
ACCESSION CA945402  
VERSION CA945402.1 GI:27433882  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
JOURNAL NIH-MGC http://mgi.nci.nih.gov/  
Unpublished National Institutes of Health, Mammalian Gene Collection (MGC)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rcmail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov

This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

FEATURES  
source

Seq primer: pYX-5.

Location/Qualifiers  
1. 687  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE: 6828925"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_id="NIH BMAP F00"  
/note="Organ: brain; Vector: pYX-Abs; Site 1: Ecor I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with Ecor I adaptor, digested with Not I, and then  
cloned directionally into pYX-Abs vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TGAGAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT 117 a 228 c 250 g 91 t 1 others  
ORIGIN

Query Match 9.1%; Score 133.4; DB 14; Length 687;  
Best Local Similarity 83.1%; Pred. No. 5.1e-22;  
Matches 152; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

705 AGCCGGCGCAAGAGGCAAGCGAGGAGCGCAACCGCATGCAACCTTAACCTCGG 764  
422 ACCCGAGGCTCAAGGCAACCAACCGAGGCAACCGCATGCAACCTTAACCTCGG 481  
765 CTGATGCGCTGCGCGGTGTCCTGCCCACTTCCCGATGACGCCAAACTTAAGAATC 824  
482 CTGAGACGCGCTGCGCGGTGTCCTGCCCACTTCCCGATGACGCCAAACTTAAGAATC 541  
825 GAGACCTGCGCTGCGCGGTGTCCTGCCCACTTCCCGATGACGCCAAACTTAAGAATC 884  
542 GAGACCTGCGCTGCGCGGTGTCCTGCCCACTTCCCGATGACGCCAAACTTAAGAATC 601  
885 GAC 887  
602 GAC 604

RESULT 5  
AL540071 1001 bp mRNA linear EST 31-MAY-2003  
LOCUS  
DEFINITION AU540071 Homo sapiens FETTL BRAIN Homo sapiens cDNA clone  
IMAGE: CS0DF035YD19 5-PRIME, mRNA sequence.  
ACCESSION AL540071  
VERSION AL540071.2 GI:31264632  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12869886.

Contract: Genoscope Centre National de Sequencage  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
2626.r. Contact: Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID: CS0DF035CB100PL.

# FEATURES

source

1. 1001  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF035YD19"  
/tissue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."  
BASE COUNT 190 a 322 c 342 g 121 t 26 others  
ORIGIN

Query Match 9.1%; Score 133; DB 9; Length 1001;  
Best Local Similarity 79.9%; Pred. No. 7.5e-22;  
Matches 151; Conservative 4; Mismatches 34; Indels 0; Gaps 0;

QY 705 AGCGGCGCAAGAGCCCAACGAGCGGAGCCCAACCGATGACAACTTAATCTCCGG 764  
DB 706 ACCCGTAGCTAAGGCAACCAACCGGAGCAACCGATGACAACTTAACCGGGA 765  
QY 765 CTGGATGCGCTGCGGCTCTCCGCACTTCCGATGAGCCCAACTTAACAAGATC 824  
DB 766 CTGGAGCGCTCGCGAGAGTGTCTCCCACTTCCGAGAGAGCCCAAGTCAACAAGATC 825  
QY 825 GAGACCTGCGCTTGGCCCAACTTATTTGGGCACTGACTCAGACGCTGGCATAGCG 884  
DB 826 GAGACGCTGCGCTTGGCCCAACTTATTTGGGCACTGACTCAGACGCTGGCATAGCG 885  
QY 885 GACCAACGC 893  
DB 886 GRTMAYTGC 894

RESULT 6  
LOCUS BJ010277 632 bp mRNA linear EST 05-DEC-2001  
DEFINITION BJ010277 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA142H01 5',  
mRNA sequence.  
ACCESSION BJ010277  
VERSION BJ010277.1 GI:117364159  
KEYWORDS EST.  
SOURCE Oryzias latipes (Japanese medaka)  
ORGANISM Oryzias latipes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;  
Belontiiformes; Adriantichthyidae; Oryziinae; Oryzias.  
REFERENCE 1 (bases 1 to 632)  
AUTHORS Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jinbo, T. and Takeda, H.  
TITLE Medaka EST Project in Takeda's lab  
JOURNAL Unpublished  
COMMENT Contact: Tadao Shin-I  
Center for Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856

FEATURES  
source  
1. 632  
/organism="Oryzias latipes"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MF01SSA142H01"  
/sex="mixture of female and male"  
/tissue\_type="whole embryo"  
/dev\_stage="segmentation stage 20 - 25"  
/clone\_lib="MF01SSA cDNA"

BASE COUNT 120 a 231 c 190 g 90 t 1 others  
ORIGIN

Query Match 8.9%; Score 130.2; DB 12; Length 632;  
Best Local Similarity 72.7%; Pred. No. 3e-21;  
Matches 168; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 658 GCGGCAACAGCCCAAGAGCGAGTTGGCACTGACGACGAGCAAGCCGCGCAAGA 717  
DB 248 GCGGAGCCGCGCGCGCTGCGAGCGGCTGCGAGGTGTGTAAGAAAGACCGCGCTGA 307  
QY 718 AGGCCAAGACCGGAGCGCAACCGCATGACCAACTTAATCTCCGCTGGATCGCTGC 777  
DB 308 AGGCTAAGACCGGAGCGCAACCGCATGACCAACTTAATCTCCGCTGGATGAGCTGC 367  
QY 778 GCGGTGTCTGCGCACTTCCCGGATGACGCGCAACTTAAGATGAGACCTCGGCT 837  
DB 368 GCGGCTGCTGCGCGCTTCCGAGCAAGCAAGTGAACCAAGATGAGATGAGCTGCGCT 427  
QY 838 TCGCCCAACTTATTTGGGCACTGACTGAGCGCTGCGCATAGCGGACC 888  
DB 428 TTGGCAACTTATCTGGGCTCTGTCGAGACCATCCGATCGCAGACC 478

RESULT 7  
LOCUS BG808248 600 bp mRNA linear EST 20-DEC-2001  
DEFINITION BG808248 2083-52 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,  
mRNA sequence.  
ACCESSION BG808248  
VERSION BG808248.1 GI:17955225  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 600)  
AUTHORS White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.  
TITLE Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)  
MEDLINE 21671825  
PUBMED 11812828

COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

FEATURES  
source  
1. 600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II library"

BASE COUNT 94 a 238 c 161 g 107 t  
ORIGIN



Query Match 8.8%; Score 128.6; DB 12; Length 600;  
 Best Local Similarity 79.6%; Pred. No. 7.3e-21;  
 Matches 152; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 729 CGGAGCGCCACCGGATGACAACTTAACCTCCGGCTGGAGTCCGCTCGCGGTCTCTG 788  
 Db 1 CGCGAGCGCCACCGGATGACAACTTAACCGCGGCTGGAGTCCGCTCGCGAGGTGTG 60  
 QY 789 CCCACCTTCGCGAGTGGAGCCCAACTTCAAGATCGAGACCCCTGCGCTTGGCCACAAAC 848  
 Db 61 CCCACCTTCGCGAGTGGAGCCCAACTTCAAGATCGAGACCCCTGCGCTTGGCCACAAAT 120  
 QY 849 TACATTTGGGACACTGACTCAGACGCTGGCGCATAGCGGACACAGCTTCTCAGGCCCGAG 908  
 Db 121 TACATCTGGGGGCTCAGCGAGACTGTGGCGCTGGCGGACACTGTGGCGCGCGCGGTGGC 180  
 QY 909 CCCCTGTGGCC 919  
 Db 181 CTCGAGGGGGC 191

RESULT 8 730 bp mRNA linear EST 20-FEB-2003  
 BUE12495  
 LOCUS UI-M-FRO-cbc-k-21-0-UI.r1 NIH BMAP\_FRO Mus musculus cDNA clone  
 DEFINITION UI-M-FRO-cbc-k-21-0-UI 5', mRNA sequence.  
 ACCESSION BUE12495  
 VERSION BUE12495.1 GI:223278710  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 730)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Clone distribution information can be obtained  
 from Dr. M. Bento Soares, bento-soares@uiowa.edu  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: PYX-5.

FEATURES  
 SOURCE Location/Qualifiers

1..730  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-FRO-cbc-k-21-0-UI"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP\_FRO"  
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1; Ecot I;  
 Site 2: Not I; The library was constructed according  
 to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AGCGAGACG. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National

Query Match 8.8%; Score 128.6; DB 13; Length 730;  
 Best Local Similarity 81.4%; Pred. No. 7.9e-21;  
 Matches 149; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 705 AGCGGCGCGCAAGAGGCGCAAGCGGAGGCGCAACCGCATGCAACCTTAACCTCGCG 764  
 Db 543 ACCGCGAGCTCAAGGCGCAACCGGAGGCGCAACCGCATGCAACCTTAAGCGCG 602  
 QY 765 CTGATGGCGCTGCGCGGTGTCTCTCCCACTTCCCGATGACGCGCAACTTAAGAATC 824  
 Db 603 CTGAGCGGCTGCGCGAGGTCTCTCCCACTTCCCGAGATGCAAGCTCAGCAAGATC 662  
 QY 825 GAGACCTGCGCTTGGCCACCACTACTTTGGGACCTGACTCAGACGCTGCGCATACG 884  
 Db 663 GAGACGCTGCGCTTGGCCACCACTACTTTGGGACCTGACTCAGACGCTGCGCATACG 722  
 QY 885 GAC 887  
 Db 723 GAC 725

RESULT 9 823 bp mRNA linear EST 26-AUG-2002  
 BUE54481  
 LOCUS UI-M-FD0-bzj-i-24-0-UI.r1 NIH BMAP\_FD0 Mus musculus cDNA clone  
 DEFINITION UI-M-FD0-bzj-i-24-0-UI 5', mRNA sequence.  
 ACCESSION BUE54481  
 VERSION BUE54481.1 GI:22494558  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 823)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

Seq primer: PYX-5.

FEATURES  
 SOURCE Location/Qualifiers

1..823  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6404447"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 12.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP\_FD0"  
 /note="Organ: brain; Vector: pYX-Asc; Site: 1; Ecot I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then

cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 155 a 273 c 289 g 104 t 2 others

Query Match 8.7%; Score 127.4; DB 13; Length 823;

Best Local Similarity 81.6%; Pred. No. 1.6e-20;

Matches 146; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 705 AGCCGCGCCAGAAAGCCCAAGCAGCGGAGCGCAACCGATGACCACTTAACCTCCGG 764  
 DB 645 ACCCGAGGCTCAAGGCCCAACACCGCAGCGACCGATGACCACTTAACCGCGG 704  
 QY 765 CTGATGCGCTGCGCGGTCTCTGCCCACTTCCCGATGACCGCAACTTACAAAGATC 824  
 DB 705 CTGAGCGGCTGCGCGAGGTGTGCTGCCCACTTCCCGAGATGCCANGCTACGANGATC 764  
 QY 825 GAGACCTGCGCTTGGCCCAACACTATTGGGCACTGACTCAGACGCTGGCATACG 883  
 DB 765 GAGACGCTGCGCTTGGCCCAACACTATTGGGCACTGACTCAGACGCTGGCATACG 823

## RESULT 10

EX419330 947 bp mRNA linear EST 15-MAY-2003

LOCUS BX419330 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone

DEFINITION CS0DF015Y006 5-PRIME, mRNA sequence.

ACCESSION BX419330

VERSION BX419330.1 GI:30765873

KEYWORDS EST.

ORGANISM Homo sapiens (human)

SOURCE Homo sapiens

COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 947)

Full-length cDNA libraries and normalization

Unpublished

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr; Web: www.genoscope.cns.fr

Library was constructed by life technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2626.x

Contact: Feng Liang Email: fliang@life-tech.com URL:

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paraday Avenue Genoscope sequence ID: CS0DF015DB03QPL

Location/Qualifiers

1..947

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0DF015Y006"

/tissue\_type="FETAL BRAIN"

/dev\_stage="fetal"

/clone\_lib="Homo sapiens FETAL BRAIN"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-strand cDNA was digested with Not I and

cloned into the Not I and EcoRV sites of the pCMVSPORT 6

vector. Library was not normalized."

BASE COUNT 73 a 393 c 315 g 130 t 36 others

ORIGIN

Query Match 8.3%; Score 121.8; DB 13; Length 947;

Best Local Similarity 72.7%; Pred. No. 4.1e-19;

Matches 152; Conservative 12; Mismatches 44; Indels 1; Gaps 1;

QY 704 AAGCGGCGCAAGAGCCCAAGCAGCGGAGCGCAACCGATGACCACTTAACCTCCGC 763  
 DB 717 AACCGTGAAGTGAAGGCCAAMAAACGAGAGCAACCGATGACCACTTAACCGCGC 776

QY 764 GCTGATGCGCTGCGCGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTACAAAGAT 823

DB 777 ACTGAGACCGCTGCGCGAGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTACAAAGAT 836

QY 824 CGAGACCTGCGCTTGGCCCAACACTATTGGGCACTGACTCAGACGCTGGCATACG 883

DB 837 MGAGAMCTGGGCTTCCG-CGMAAAMAAAMWTGGGCACTACCGARACCTCGCGCTGGC 895

QY 884 GAGACCAAGCTTCTACGCGCCCGAGCCG 912

DB 896 GATCACTGCGGCGGCGCGCGGCGGCGG 924

RESULT 11

BU924937 600 bp mRNA linear EST 30-OCT-2002

LOCUS BU924937 7103-91 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,

DEFINITION mRNA sequence.

ACCESSION BU924937

VERSION BU924937.1 GI:24428820

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 600)

Mu.X., Zhao,S., Pershad,R., Hejeh,T.-F., Scarpa,A., Wang,S.W.,

White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.

Gene expression in the developing mouse retina by EST sequencing

and microarray analysis

Nucleic Acids Res. 29 (24), 4983-4993 (2001)

Contact: Klein WH

Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center

Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

Tel: 713 792 3646

Fax: 713 790 0329.

Location/Qualifiers

1..600

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/tissue\_type="neural retina"

/dev\_stage="embryonic day 14.5 post-fertilization"

/clone\_lib="Mouse E14.5 retina lambda ZAP II library"

BASE COUNT 99 a 194 c 222 g 82 t 3 others

ORIGIN

Query Match 8.2%; Score 120.4; DB 13; Length 600;

Best Local Similarity 75.1%; Pred. No. 7.5e-19;

Matches 163; Conservative 0; Mismatches 52; Indels 2; Gaps 1;

QY 705 AGCCGCGCCAGAAAGCCCAAGCAGCGGAGCGCAACCGATGACCAACTTAACCTCCGC 764

DB 336 ACCCGAGGCTCAAGAGCCCAACACCGGAGCGCAACCGATGACCAACTTAACCGCGC 395

QY 765 CTGATGCGCTGCGCGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTACAAAGATC 824

DB 396 CTGAGCGGCTGCGCGAGGTGTCTCTGCCCACTTCCCGAGATGCCAACTCAGAAAGATC 455

QY 825 GAGACCT--GCGCTTCCCGACCAACTATTGGGCACTGACTCAGACGCTGGCATAG 882

DB 456 GAGACGCTTGGCGCTTCCCGCCCAACTTACACTTGGGCTACCGAAGCTTGGCGCTGG 515

QY 883 CGACACAGAGTTTACAGCGCCCGAGCGCCCTGTGCC 919

DB 516 CGACCACTGCGCGCGCGCGGTGGCTTTCAGAGGCGC 552



Db 237 CGCATGGCCGATCA 224

RESULT 14	LOCUS	401 bp	mRNA	linear	EST 19-FEB-2003
AM147434	AM147434				
DEFINITION	daa2h12.y1 Xenopus laevis oocyte Xenopus laevis cDNA clone				

ACCESSION	AM147434
VERSION	AM147434.1
KEYWORDS	GI:6195330
SOURCE	EST.
ORGANISM	Xenopus laevis (African clawed frog)
	Xenopus laevis

REFERENCE  
AUTHORS  
1 (bases 1 to 401)  
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Page, D.

TITLE	WashU Xenopus EST project, 1999
JOURNAL	unpublished
COMMENT	Contact: Sandy Clifton, Ph.D.

FEATURES	Location/Qualifiers
source	1. .401

BASE COUNT	87 a	140 c	84 g	90 t
ORIGIN				

Query Match	7.6%;	Score 110.8;	DB 9;	Length 401;
Best Local Similarity	74.7%;	Pred. No. 14e-16;		
Matches 139;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0

Oy 704 AAGCGGCGCAGAAAGGCCAACGACCGGAGGCGCAACCGCATCAGCAACTTAACTCCG 763  
 Db 4 AACC CGCGCGCTTAAAGCCAAATTAACCGCGAGAGGAAATGCAATCCACCACTGAACCTATGC 63

0y 764 GCTGGATGCGCTGCGCGGTGTCTGCGCCACTTCCCGGATACGCGCAACTTACAAAGAT 823

Db 64 GCTCGATTCTGAGGAGGTCCTACCGTCAATTACCGAAGACGCCAACTCACCAAGAT 123

824 CGAGACCCCTGCGCTCGCCACACATCATTTGGGACTGCTAGCGCTGGCATAGC 883  
 |||||  
 124 AGAGACCTTGGCGCTTTGCCACAATACATCTGGGCTTTAGCGAAACTTTGGCCCTGGC 183

QY	884	GGACCA	885
Db	184	CGACCA	185

RESULTS	15
CD282259	
LOCUS	
DEFINITION	
G33271..16 NCI_CGAP_ZemBd	603 bp mRNA linear EST 23-MAY-2003
mRNA sequence.	
CD282259	

ACCESSION	C4262235
VERSION	CD282259.1
KEYWORDS	EST.
SOURCE	Danio rerio (zebrafish).
ORGANISM	Danio rerio
	Database: Metazoa; Chromosome: 1

.....	AUTHORS	Amundsen, C., Cachuela, N., Chen, F., Cheung, L. M., Chong, A., Murray, L., Oliva, J., Park, C., Reyes, J., Yungen, J. and Swimmer, C.
.....	TITLE	Expressed sequence tags from NCI_CGAP_Zemba2, a Dantio retilo embryonic library
.....	JOURNAL	Unpublished
.....	COMMENT	Contact: Chen F.

FEATURES	Location/Qualifiers
source	1. .603

BASE COUNT	159 a	185 c	155 g	104 t
ORIGIN				

Query Match	7.5%;	Score 110.2;	DB 14;	Length 603;
Best Local Similarity	74.3%;	Pred. No. 2.4e-16;		
Matches 139;	Conservative	0;	Mismatches 48;	Indels 0;
				Gaps 0

Df  
Dy  
705  
AGCGGCGCAGAAAGGC CAAAGACCCGGAGAGCACACCAATCTTACCACCGG  
| | | | | | | | | | | | | | | | | | | | | |  
325 AACCGAGGCTGAAGGCCAACGACCCGAGAGGAACAAGATGCACAACTTAACAGCGA 38

QY 765 CTGATATCGCTGGCGCGATGTCCTGAGCCAACTTCCTCGGATGAGCGCCAACTTACCAAGTC 82

Db 385 TTGGATCTTTGAGAAAGCGCTCGCTGCTTCTCGAGACACAAAGCTGACCAAAATT 44

QY 825 GAGACCTCGCGCTTGGCCCAACAATATTGGGCACTGACTCGAGACGCTGCGCATACG 88

Db 445 GAGACTCTGGCGCTCGCTACACTCTGSGCACTTTCGAGACCATCCGAATGCA 504  
 Oy 885 GACCAACA 891  
 Db 505 GACCAAGA 511

Search completed: January 26, 2004, 22:05:05  
Job time : 3312 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 18:54:03 ; Search time 5306 Seconds

(without alignments)  
11256.701 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcagtagcagagagagcag.....agagtcacccaatccagtcgt 1460

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

```
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1409	96.5	1491	6 A91167	A91167 Sequence 1
2	1409	96.5	1491	6 BD023626	BD023626 Polypepti
3	1409	96.5	1491	10 RNRELAXT	Y10619 R.norvegicu
4	1060	72.6	258815	2 AC127817	AC127817 Rattus no
5	92	6.3	861	6 AX698801	AX698801 Sequence
6	92	6.3	861	10 MMU76208	U76208 Mus musculu
7	92	6.3	1861	10 AF364300	AF364300 Mus muscu
8	92	6.3	5567	10 MMATH4B	Y09167 M.musculus
9	92	6.3	138070	2 AC109783	AC109783 Mus muscu
10	92	6.3	215050	2 AC127417	AC127417 Mus muscu
11	35	2.4	2370	9 BC036847	BC036847 Homo sapi
12	35	2.4	6123	9 AF303002	AF303002 Homo sapi
13	35	2.4	91531	2 AC079846_3	Continuation (4 of
14	35	2.4	179697	9 AC023886	AC023886 Homo sapi
15	32	2.2	1330	9 HSA133776	A113776 Homo sapi
16	32	2.2	5340	9 AF234829	AF234829 Homo sapi
17	32	2.2	165110	9 AL450311	AL450311 Human DNA
18	32	2.2	173341	2 AC021954	AC021954 Homo sapi
19	26	1.8	790	6 AX548297	AX548297 Sequence
20	26	1.8	170896	2 AC011010	AC011010 Homo sapi
21	25	1.7	25	6 A91170	A91170 Sequence 4
22	25	1.7	25	6 BD023629	BD023629 Polypepti
23	23	1.6	735	10 MMU67776	U67776 Mus musculu
24	23	1.6	738	6 AR308548	AR308548 Sequence
25	23	1.6	770	5 AF123884	AF123884 Gallus ga
26	23	1.6	790	5 GGA012659	AJ012659 Gallus ga
27	23	1.6	932	10 MMU63841	Y09166 M.musculus
28	23	1.6	1315	6 MMU63841	U63841 Mus musculu
29	23	1.6	1333	6 AR023715	AR023715 Sequence
30	23	1.6	1333	6 AR225848	AR225848 Sequence
31	23	1.6	1341	5 AF109014	AF109014 Gallus ga
32	23	1.6	1385	10 MMU76207	U76207 Mus musculu
33	23	1.6	1412	10 MMATH4A	Y07621 M.musculus
34	23	1.6	1880	5 AF303000	AF303000 Gallus ga
35	23	1.6	10393	10 AF303001	AF303001 Mus muscu
36	23	1.6	71538	2 AC118243	AC118243 Mus muscu
37	23	1.6	123855	2 AC102600	AC102600 Mus muscu
38	23	1.6	149268	2 AC124395	AC124395 Mus muscu
39	23	1.6	262798	2 AC111702	AC111702 Rattus no
40	22	1.5	215591	2 AC102159	AC102159 Mus muscu
41	21	1.4	21	6 AX553597	AX553597 Sequence
42	21	1.4	21	6 AX553645	AX553645 Sequence
43	21	1.4	310	6 BC052235	BC052235 Homo sapi
44	21	1.4	1058	9 BC052235	BC052235 Homo sapi
45	21	1.4	1164	5 GENEURODL	Y09596 G.gallus mr

## ALIGNMENTS

RESULT 1  
LOCUS A91167 1491 bp DNA linear PAT 22-JAN-2000  
DEFINITION Sequence 1 from Patent WO9827206.  
ACCESSION A91167  
VERSION A91167.1 GI:6740202  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus sp.  
Rattus sp.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 1491)  
Icard-Liepkains, C., Maliet, J. and Corresponding, N.A.  
Patent: WO 9827206-A 1 25-JUN-1998;

REFERENCE  
AUTHORS: Icard-Liepkains, C., Maliet, J. and Corresponding, N.A.  
JOURNAL: Patent: WO 9827206-A 1 25-JUN-1998;





PR	19-DEC-1996	FR	96/15651	
PI	CHRISTINE ICARD	LIEPKALNS, JACQUES	MALLET, PHILIPPE	RAVASSARD P
CO	7K14/47, A6IK31/711, A6IK35/76, A6IK38/00, A6IK48/00, A6IP25/00, PC			
C1	2N15/09,			
PC	C12N15/00, A6IK37/02			
CC	Strandedness: Single;			
CC	Topology: Linear;			
PH	Key	Location/Qualifiers.		
FEATURES				
source		1..1491		
		/organism="Rattus sp."		
		/mol_type="Genomic DNA"		
		/db_xref="taxon:10118"		
BASE COUNT	307 a	487 c	413 g	284 t
ORIGIN				
Query Match	96.5%;	Score 1409;	DB 6;	Length 1491;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1459;	Conservative	0;	Mismatches	1;
			Indels	0;
			Gaps	0
OY	1	GCAGGTAGCGAGAGAGCAGTCCTCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAAGCA	60	
Db	1	GCAGGTAGCGAGAGAGAGCAGTCCTCTGGGCCCCCGTTGCTGATTGGCCCGTGGCACAAGCA	60	
OY	61	GCAGCCCCGGCAGGACACGCTCCCTGGTCCCGGGCAGACAGATAAAGCTGCGACAGGACACA	120	
Db	61	GCAGCCCCGGCAGGACACGCTCCCTGGTCCCGGGCAGACAGATAAAGCTGCGACAGGACACA	120	
OY	121	CGATTAGAGCTCAAGAGTCCCTCGTGGGTCTACCACTTSCACAGAGGCGGAGACCCCT	180	
Db	121	CGATTAGAGCTCAAGAGTCCCTCGTGGGTCTACCACTTSCACAGAGGCGGAGACCCCT	180	
OY	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCAGGCGAGGGCGGCTTGACCTCAG	240	
Db	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCAGGCGAGGGCGGCTTGACCTCAG	240	
OY	241	CAAACTTCGAAGCGAGCAGAGAGGGTTACGCTATCCACCGCTTGACTTGACACACC	300	
Db	241	CAAACTTCGAAGCGAGCAGAGAGGGTTACGCTATCCACCGCTTGACTTGACACACC	300	
OY	301	GCAGCTCTCTGTTCTTTTGGGCCGGAGATPACTAGGTAACTTTAGAACTCCAAAGGG	360	
Db	301	GCAGCTCTCTGTTCTTTTGGGCCGGAGATPACTAGGTAACTTTAGAACTCCAAAGGG	360	
OY	361	TAGAAAGAGGAGTGGGTGGGCGTACTCTTAGTCCGCGTGAAGTACCTCTAAGTCAGAG	420	
Db	361	TAGAAAGAGGAGTGGGTGGGCGTACTCTTAGTCCGCGTGAAGTACCTCTAAGTCAGAG	420	
OY	421	ACTGTCAACACCCCTCTTCAATTTTCCCACTTGAAGATGCGGCTCATCTCCCTTGATG	480	
Db	421	ACTGTCAACACCCCTCTTCAATTTTCCCACTTGAAGATGCGGCTCATCTCCCTTGATG	480	
OY	481	CGCCCAACATCCAAAGTGTCCAAAGAGACCCACAGACCCCTTCCCGGAGCTCGAGCCAG	540	
Db	481	CGCCCAACATCCAAAGTGTGTCCAAAGAGACCCACAGACCCCTTCCCGGAGCTTGAGCCAG	540	
OY	541	AAGTGCTCAGTTCCAATTCACCCCACTTAGGCCCACTTCTGTAACGAGGAGCTGCTCG	600	
Db	541	AAGTGCTCAGTTCCAATTCACCCCACTTAGGCCCACTTCTGTAACGAGGAGCTGCTCG	600	
OY	601	AAAGCAAGAGCAGGTGATCTGCGAGGAGCATGAGAGAAAGTCCGTGCGCGCGCGAGAGGC	660	
Db	601	AAAGCAAGAGCAGGTGATCTGCGAGGAGCATTCAGAGAAAGTCCGTGCGCGCGCGAGAGGC	660	
OY	661	GCAAGAGGCGCCAAAGGCGAGTGGGACTGAGAGAGAGGAGCAAGCGCGGCGCAAGAAAG	720	
Db	661	GCAAGAGGCGCCAAAGAGCGAGTGGGACTGAGAGAGAGGAGCAAGCGCGGCGCAAGAAAG	720	
OY	721	CCAAAGCAGCCGGAGGCGCAACCGCATGCAACAACTTTAACTCCGCGTGAATGCGCTGCGG	780	
Db	721	CCAAAGCAGCCGGAGGCGCAACCGCATGCAACAACTTTAACTCCGCGTGAATGCGCTGCGG	780	
OY	781	GTGTCTCTGCCACCTTCCCGGATGACGCCCAAACTTACAAAGATCGAAGACCTTGCGCTTCG	840	

Db	781	GTGTCTCTGCCCACCTTCCCGATGAGCCCAACTTACAAAGATGAGACCTTGGCCTTGC	840
QY	841	CCCAACAATCATTTGGGCACTGACTCAGACGCTGCGCATGACGAGACCAACAGCTTCTACG	900
Db	841	CCCAACAATCATTTGGGCACTGACTCAGACGCTGCGCATGACGAGACCAACAGCTTCTACG	900
QY	901	GCCTCGAGCCCTCTGTGCTCTGTGGGAGCTGGGAGACCGCGGAGAGGGGGCTTCAGCGGC	960
Db	901	GCCTCGAGCCCTCTGTGCTCTGTGGGAGCTGGGAGACCGCGGAGAGGGGGCTTCAGCGGC	960
QY	961	ACTGGGGGCTCATCTACCTCCAGTTTCCAAAGCGGTGATGAGGCTGAGCCCAAGCTCAT	1020
Db	961	ACTGGGGGCTCATCTACCTCCAGTTTCCAAAGCGGTGATGAGGCTGAGCCCAAGCTCAT	1020
QY	1021	TGAGAGGATTCCTGCGCTGCGAGGTGCGCACTCCCAATCCCTGTCTGCTCCCGGACACC	1080
Db	1021	TGAGAGGATTCCTGCGCTGCGAGGTGCGCACTCCCAATCCCTGTCTGCTCCCGGACACC	1080
QY	1081	TGCTGTCTCAGACTTCTTGTGAAGGGGCCAAACAGAGCTTGGGCGCTGGCAG	1140
Db	1081	TGCTGTCTCAGACTTCTTGTGAAGGGGCCAAACAGAGCTTGGGCGCTGGCAG	1140
QY	1141	AAAGGAGGAGGTGACAGAGCTGCTGAATGGAATGGAAGGTATGAGGCACTCGAGCATCTGC	1200
Db	1141	AAAGGAGGAGGTGACAGAGCTGCTGAATGGAATGGAAGGTATGAGGCACTCGAGCATCTGC	1200
QY	1201	CCCTTCTGGCTTTCATTAAGTCAGAGTCCCTGATTTTAAACAGGATTCGACAGTTCCTTCT	1260
Db	1201	CCCTTCTGGCTTTCATTAAGTCAGAGTCCCTGATTTTAAACAGGATTCGACAGTTCCTTCT	1260
QY	1261	GCTGTGCGTCGACAAAGATATTTGACGGCTGATCTCTTTAACCCCTCTCAGTGTGGCC	1320
Db	1261	GCTGTGCGTCGACAAAGATATTTGACGGCTGATCTCTTTAACCCCTCTCAGTGTGGCC	1320
QY	1321	ACCTCAAACTCCCGCTCCAAAGAGAGGACCCGTAAGCATTAATATGTTGGAGACTGCC	1380
Db	1321	ACCTCAAACTCCCGCTCCAAAGAGAGGACCCGTAAGCATTAATATGTTGGAGACTGCC	1380
QY	1381	ATACTCTCTGATGACTCGCCCTCTTTCAATCTGCGGCTTCCACACACCGCTTCTCC	1440
Db	1381	ATACTCTCTGATGACTCGCCCTCTTTCAATCTGCGGCTTCCACACACCGCTTCTCC	1440
QY	1441	AGAGTGACCTATCCAGTGT	1460
Db	1441	AGAGTGACCTATCCAGTGT	1460
RESULT 3			
RNRELAXT			
LOCUS	RNRELAXT	1491 bp	linear
DEFINITION	R.norvegicus mRNA for transcriptional regulator, Relax.		ROD 06-MAY-1997
ACCESSION	Y10619		
VERSION	Y10619.1	GI:2072737	
KEYWORDS	Relax; transcriptional regulator.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1		
AUTHORS	Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.		
TITLE	Relax, a novel rat bHLH transcriptional regulator transiently expressed in the ventricular proliferating zone of the developing central nervous system		
JOURNAL	J. Neurosci. Res. 48 (2), 146-158 (1997)		
MEDLINE	97276390		
PUBMED	9130143		
REFERENCE	2	(bases 1 to 1491)	
AUTHORS	Ravassard, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JUN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI, Hôpital de la Plie Salpêtrière, 83 Bd. de l'Hôpital, F-75013		

FEATURES  
source  
Paris, FRANCE  
Location/Qualifiers  
1.1491  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Wistar"  
/db\_xref="taxon:10116"  
/clone\_1ib="RT-PCR"  
459..1103  
/function="transcriptional regulator"  
/codon\_start=1  
/product="Relax"  
/protein\_id="CAA71630.1"  
/db\_xref="GI:2072738"  
/db\_xref="SPTREMBL:O08718"  
/translation="MAPK14P1TOVSETOQPPFGASDHEVLSNSTPSPITVPR  
DSEIAGDCKRSTRLKARSGRNPSELSIKORRKRKKANDRENNHNLISA  
IDALRGVLPFPDDAKLTKETLRFANHYIMALTORLADHSFYGPPEPVPCEIGS  
PGSGSGDMGSIYSPVQAGSISPTASLEPPLQVPSPSCLIPQLVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN

Query Match 96.5%; Score 1409; DB 10; Length 1491;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GCAGGTAGGAGAGGAGGAGTCCCTGGGCCCCGTTGATTGGCCGTCGACAGGGA 60  
1 GCAGGTAGGAGAGGAGGAGTCCCTGGGCCCCGTTGATTGGCCGTCGACAGGGA 60  
61 GCAGCCCGGACAGGAGCTCTGGTCCGGGACAGAGAGATAAGCTGACAGGAGACA 120  
61 GCAGCCCGGACAGGAGCTCTGGTCCGGGACAGAGAGATAAGCTGACAGGAGACA 120  
61 GCAGCCCGGACAGGAGCTCTGGTCCGGGACAGAGAGATAAGCTGACAGGAGACA 120  
121 CGATTAGAGCTCAGAACTCTCTGAGTCTCACTGACAGAGAGCCGACCTT 180  
121 CGATTAGAGCTCAGAACTCTCTGAGTCTCACTGACAGAGAGCCGACCTT 180  
121 CGATTAGAGCTCAGAACTCTCTGAGTCTCACTGACAGAGAGCCGACCTT 180  
181 CGAGGTTCTTGTCTCTCTCCAGAGCAATTACTCAGAGAGAGCCGCTGAGTCA 240  
181 CGAGGTTCTTGTCTCTCTCCAGAGCAATTACTCAGAGAGAGCCGCTGAGTCA 240  
181 CGAGGTTCTTGTCTCTCTCCAGAGCAATTACTCAGAGAGAGCCGCTGAGTCA 240  
241 CAAAATTGAGAGGAGAGAGAGGTTCACTACAGCTCTTGAACCTGACACACC 300  
241 CAAAATTGAGAGGAGAGAGGTTCACTACAGCTCTTGAACCTGACACACC 300  
301 GCAGCTCTCTGTTCTTTTGAAGCCCGAGTAAGTAACTTTAGGAACCTTCAAGG 360  
301 GCAGCTCTCTGTTCTTTTGAAGCCCGAGTAAGTAACTTTAGGAACCTTCAAGG 360  
301 GCAGCTCTCTGTTCTTTTGAAGCCCGAGTAAGTAACTTTAGGAACCTTCAAGG 360  
361 TAGAAGAGGAGATGGGTGGGCTACTCTAGTCCCGCTGGAGTGAACCTTAGTGAG 420  
361 TAGAAGAGGAGATGGGTGGGCTACTCTAGTCCCGCTGGAGTGAACCTTAGTGAG 420  
361 TAGAAGAGGAGATGGGTGGGCTACTCTAGTCCCGCTGGAGTGAACCTTAGTGAG 420  
421 ACTGTCAACCCCTTCCATTTTCCAACTCAGAGATGGGCTCACTCCCTGGAGT 480  
421 ACTGTCAACCCCTTCCATTTTCCAACTCAGAGATGGGCTCACTCCCTGGAGT 480  
421 ACTGTCAACCCCTTCCATTTTCCAACTCAGAGATGGGCTCACTCCCTGGAGT 480  
481 CGGCCACCATCCAAAGTGTCCCAAGAGACCCAGAACCTTTCCGAGAGCTCGACAC 540  
481 CGGCCACCATCCAAAGTGTCCCAAGAGACCCAGAACCTTTCCGAGAGCTCGACAC 540  
481 CGGCCACCATCCAAAGTGTCCCAAGAGACCCAGAACCTTTCCGAGAGCTCGACAC 540  
541 AATGTCTCAATTCCATTTCCACCCCTAGGCCCACTCTGTAACGAGGACCTGCTCC 600  
541 AATGTCTCAATTCCATTTCCACCCCTAGGCCCACTCTGTAACGAGGACCTGCTCC 600  
541 AATGTCTCAATTCCATTTCCACCCCTAGGCCCACTCTGTAACGAGGACCTGCTCC 600  
601 AACGAGAGAGGAGTGTCCGAGAGGACATCGAGGAAGCTCCGTCGCGCGGAGAGGC 660  
601 AACGAGAGAGGAGTGTCCGAGAGGACATCGAGGAAGCTCCGTCGCGCGGAGAGGC 660  
601 AACGAGAGAGGAGTGTCCGAGAGGACATCGAGGAAGCTCCGTCGCGCGGAGAGGC 660  
661 GCAACAGGAGGAGGAGGAGTGTGCACTGAGCAAGAGAGAGAGGAGGAGAGG 720  
661 GCAACAGGAGGAGGAGGAGTGTGCACTGAGCAAGAGAGAGAGGAGAGGAGAGG 720

721 CCAACGACCGGAGAGGAGCAACCGCATGCAAACTTAATCCGCGCTGAGTGCCTGCG 780  
721 CCAACGACCGGAGAGGAGCAACCGCATGCAAACTTAATCCGCGCTGAGTGCCTGCG 780  
781 GTGTCTTCCGACCTTCCCGAGTGAAGCCCAAACTTAAAGATGAGACCTTGCCTTCC 840  
781 GTGTCTTCCGACCTTCCCGAGTGAAGCCCAAACTTAAAGATGAGACCTTGCCTTCC 840  
841 CCAACAACTTAATTTGGGCACTGACTGAGAGCTGGGATGAGCCGACCAAGCTTCAAG 900  
841 CCAACAACTTAATTTGGGCACTGACTGAGAGCTGGGATGAGCCGACCAAGCTTCAAG 900  
841 CCAACAACTTAATTTGGGCACTGACTGAGAGCTGGGATGAGCCGACCAAGCTTCAAG 900  
901 GCCCGAGAGCCCTTGTGCTGTGGGAGCTGGAGAGCCCGAGAGAGGAGGCTTCAAGCG 960  
901 GCCCGAGAGCCCTTGTGCTGTGGGAGCTGGAGAGCCCGAGAGAGGAGGCTTCAAGCG 960  
961 ACTGAGGCTCTATCTACTTCCCAAGTTTCCCAAGCTGTAGCTGAGCCCAAGCTTCA 1020  
961 ACTGAGGCTCTATCTACTTCCCAAGTTTCCCAAGCTGTAGCTGAGCCCAAGCTTCA 1020  
1021 TGGAGAGTCCCTGCTGAGAGTCCCAAGCTTCCCTGCTGCTCCCGGACACCC 1080  
1021 TGGAGAGTCCCTGCTGAGAGTCCCAAGCTTCCCTGCTGCTCCCGGACACCC 1080  
1081 TGGTGTCTCAGACTTCTTGTGAAGGAGCCCAAGAGCTGGAGGCTGGGAGCTGGAG 1140  
1081 TGGTGTCTCAGACTTCTTGTGAAGGAGCCCAAGAGCTGGAGGCTGGGAGCTGGAG 1140  
1141 AAGAGAGAGGAGTCAAGCTGTCTGAATAATGAGAGTGTAGAGCATCTGACATCTG 1200  
1141 AAGAGAGAGGAGTCAAGCTGTCTGAATAATGAGAGTGTAGAGCATCTGACATCTG 1200  
1201 CCTTCTGCTTCAATTAATGAGTCCCTGATTTAAACAGAGTTTGGACAGTCTTCT 1260  
1201 CCTTCTGCTTCAATTAATGAGTCCCTGATTTAAACAGAGTTTGGACAGTCTTCT 1260  
1261 GCTGTGCGTCAACAAGGATATTGAGAGCTGATCTCTTAAACCTCTCTGAGTGGCC 1320  
1261 GCTGTGCGTCAACAAGGATATTGAGAGCTGATCTCTTAAACCTCTCTGAGTGGCC 1320  
1261 GCTGTGCGTCAACAAGGATATTGAGAGCTGATCTCTTAAACCTCTCTGAGTGGCC 1320  
1321 ACCTCAAACTCCCGCTCCAAAGAGAGAGAGCCGAGACCTAATATGTTGGAGACTGCC 1380  
1321 ACCTCAAACTCCCGCTCCAAAGAGAGAGAGCCGAGACCTAATATGTTGGAGACTGCC 1380  
1381 ATACTTCTGAGTCTCGGCTCTTTTCAAACTGCGGAGCTTCCCAACCGCTTCTCC 1440  
1381 ATACTTCTGAGTCTCGGCTCTTTTCAAACTGCGGAGCTTCCCAACCGCTTCTCC 1440  
1441 AAGTGAACCTAATCCAGTGT 1460  
1441 AAGTGAACCTAATCCAGTGT 1460

RESULT 4  
AC127817 258815 bp DNA linear HTG 19-NOV-2002  
LOCUS  
DEFINITION  
Rattus norvegicus clone CH230-259616, WORKING DRAFT SEQUENCE, 3  
unorderd pieces.  
AC127817  
AC127817.3 GI:25077905  
VERSION  
HTG: HTGS PHASE1; HTGS DRAFT; HTGS\_FULFILLTOP.  
KEYWORDS  
Rattus norvegicus (Norway rat)  
SOURCE  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 258815)  
REFERENCE  
AUTHORS  
Muzny, D., Marie, M., Metzger, M., Lee, A., Adnan, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsdorfs, S., Amin, A., Anguiano, D.,  
Anyalebech, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Bernhardt, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., d'Souza, L., Davila, M.L., Davis, C., Day-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Gantla, R., Garcia, A., Garner, T., Garza, M., Gebregioris, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Holling, B., Howles, S., Huylk, S., Hume, J., Idledit, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Kaprathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawlin, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokweme, O., Okumu, G., Olarnunagoon, A., Pal, S., Parke, K., Pasernak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Quirio, J., Rachlin, E., Reeves, K., Reiter, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Sherry, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Stead, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Woodson, H., Wolley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., Von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

**TITLE**  
Unpublished  
2 (bases 1 to 258815)  
**REFERENCE**  
Worley, K.C.  
**AUTHORS**  
Direct Submission  
**TITLE**  
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 258815)  
**AUTHORS**  
Rat Genome Sequencing Consortium.  
**TITLE**  
Direct Submission  
**JOURNAL**  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
**COMMENT**  
On Nov 19, 2002 this sequence version replaced gi:23912578.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine

Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information  
Center project name: GZXS  
Center clone name: CH230-259616  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 224747 bases at least Q40  
Consensus quality: 227981 bases at least Q30  
Consensus quality: 229752 bases at least Q20  
Estimated insert size: 228243; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbankdraft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html)).  
\* NOTE: This is a "working draft" sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 255979: contig of 255979 bp in length  
\* 255980 256079: gap of unknown length  
\* 256080 257349: contig of 1270 bp in length  
\* 257350 257449: gap of unknown length  
\* 257450 258815: contig of 1366 bp in length.  
Location/Qualifiers  
1. 258815  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-259616"  
1. 1609  
/note="wgs end extension  
clone end:sp6"  
complement(2865..3659)  
/note="clone\_boundary  
clone\_end:sp6  
site:  
end\_sequence:BZ169379"  
224098..225157  
/note="clone\_boundary  
clone\_end:T7  
site:  
end\_sequence:BZ169378"  
BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others  
ORIGIN  
Query Match 72.6%; Score 1060; DB 2; Length 258815;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 350 CCTCAAGGGTATGAAGAGGGGAGTGGGCGTACTGATCCCGGTGAGTACCT 409  
DB 247751 CCTCAAGGGTATGAAGAGGGGAGTGGGCGTACTGATCCCGGTGAGTACCT 247810  
QY 410 CTAAGTCAGAGACTGTCAACACCCCTTCATTTTTCCTCAACCTCAGATGCGCTCA 469  
DB 247811 CTAAGTCAGAGACTGTCAACACCCCTTCATTTTTCCTCAACCTCAGATGCGCTCA 247870  
QY 470 TCCCTTGATGCGCCCATCTCAAGTGTCCCAAGACCCAGACACCTTTCCCGAGC 529  
DB 247871 TCCCTTGATGCGCCCATCTCAAGTGTCCCAAGACCCAGACACCTTTCCCGAGC 247930  
QY 530 CTCGAGCAGCAAGTGTCAATTCACACCCCATCTGAGCCCATCTGATGAGCAG 589  
DB 247931 CTCGAGCAGCAAGTGTCAATTCACACCCCATCTGAGCCCATCTGATGAGCAG 247990  
QY 590 GGAAGTCTCGAAGCAGCAAGTGTGAGCAGGAGCATCTGAGAGCTCCGTGCGG 649

```

Db      247991 GGAAGTCTCCAGAGCAGAGTGAAGTCCGAGGAGCATGAGAGAGCTCCGTCGCG 248050
Qy      650 GGGCGGAGGGCGCAAGAGCCCAAGAGGAGTGGCTGAGCAAGCAGAGCAAGAGCGG 709
Db      248051 GGGCGGAGGGCGCAAGAGCCCAAGAGGAGTGGCTGAGCAAGCAGAGCAAGAGCGG 248110
Qy      710 GCGCAAGAGGCGCAAGAGCCGAGGAGCGCAAGCAGATGCAAACTTAACTCCGCGCTGGA 769
Db      248111 GCGCAAGAGGCGCAAGAGCCGAGGAGCGCAAGCAGATGCAAACTTAACTCCGCGCTGGA 248170
Qy      770 TCGCGTCCGCGGTCTCTCCCACTTCCCGATGAGCCCAAACTTAAAGATGAGAC 829
Db      248171 TCGCGTCCGCGGTCTCTCCCACTTCCCGATGAGCCCAAACTTAAAGATGAGAC 248230
Qy      830 CCGCGCTTCCGCGCAAGTCACTTGGGACAGTCAAGAGCTCCGAGTACCGGAGCGGAGCA 889
Db      248231 CCGCGCTTCCGCGCAAGTCACTTGGGACAGTCAAGAGCTCCGAGTACCGGAGCGGAGCA 248290
Qy      890 CAGCTTCTACGCGCCCGAGCCCGCTGTCCTGCGGAGCTGGAGAGCCCGGAGGAGG 949
Db      248291 CAGCTTCTACGCGCCCGAGCCCGCTGTCCTGCGGAGCTGGAGAGCCCGGAGGAGG 248350
Qy      950 CTCGAGCGGCGAGTGGGCTCTATCTACTGCCAGTTCCTCCAGCTGGTACCTGAGCCC 1009
Db      248351 CTCGAGCGGCGAGTGGGCTCTATCTACTGCCAGTTCCTCCAGCTGGTACCTGAGCCC 248410
Qy      1010 CACAGCTGATTTGAGAGAGTTCCTGCGGAGTGGAGTCCGAGTCCGAGTCCGCTGAGT 1069
Db      248411 CACAGCTGATTTGAGAGAGTTCCTGCGGAGTGGAGTCCGAGTCCGAGTCCGCTGAGT 248470
Qy      1070 CCGCGGCAAGCTGCTGCTCTGAGACTTCTTGAGAGGCGCCAAACAGGCGCTGGGCGGTG 1129
Db      248471 CCGCGGCAAGCTGCTGCTCTGAGACTTCTTGAGAGGCGCCAAACAGGCGCTGGGCGGTG 248530
Qy      1130 GGGCGCTGCGAAGAGGAGGAGTCAAGAGTGTCTGAATGAGAGTGTGAGGCACTC 1189
Db      248531 GGGCGCTGCGAAGAGGAGGAGTCAAGAGTGTCTGAATGAGAGTGTGAGGCACTC 248590
Qy      1190 GAGCATCTGCGCCCTTCTGCTTCAATGAGTCCGAGTTCCTGATTTAAACAGGATTTGAC 1249
Db      248591 GAGCATCTGCGCCCTTCTGCTTCAATGAGTCCGAGTTCCTGATTTAAACAGGATTTGAC 248650
Qy      1250 AGTTCCTGCTGCTGCTGCTGCAAGAGATTTGAGAGTGTCTCTTAAACCTCC 1309
Db      248651 AGTTCCTGCTGCTGCTGCTGCAAGAGATTTGAGAGTGTCTCTTAAACCTCC 248710
Qy      1310 TCAGTGTGGCCCACTTCAAACTCCCGCTCCAGAGAGAGAGCGGTAGCACTAAATAGTT 1369
Db      248711 TCAGTGTGGCCCACTTCAAACTCCCGCTCCAGAGAGAGAGCGGTAGCACTAAATAGTT 248770
Qy      1370 GGGAGCTCCCACTTCTCTGAGTCTCCGCTCTTAAATCTCCGCGCTCCCAACCA 1429
Db      248771 GGGAGCTCCCACTTCTCTGAGTCTCCGCTCTTAAATCTCCGCGCTCCCAACCA 248830
Qy      1430 CCGCTTCTCCAGAGTGAAGTCAAGTGT 1460
Db      248831 CCGCTTCTCCAGAGTGAAGTCAAGTGT 248861

```

```

RESULT 5
AX698801 861 bp DNA linear PAT 02-APR-2003
LOCUS AX698801
DEFINITION Sequence 7 from Patent WO02086107.
ACCESSION AX698801
VERSION AX698801.1 GI:29499589
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Wobus, A.M., St-Onge, L., Blyszczuk, P. and Hoffmann, U.
TITLE A method for differentiating stem cells into insulin-producing

```

```

JOURNAL
Patent: WO 02086107-A 7 31-OCT-2002;
Developed by Aktiengesellschaft fuer Entwicklungsbiologische Forschung
(DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
(DE)
FEATURES
source
Location/Qualifiers
1..861
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
160..804
/feature="unnamed protein product"
/codon_start=1
/protein_id="CAD87442.1"
/db_xref="GI:29499590"
/translation="MAPPRDALITQVSPFQPPGASDHEVLSSNSTPPSPTLIPR
DQSEAEVDGCTSRKLRARAGRRPSEIALSKQRSRKRRKANDRRMHNLSA
LDALGVLPDPDDKLTIRLIRAHNYIALTQIADHSFYGPFPYPCGLGS
FGGSGNGWGSISYSPVQAGNLSPTASLEFPFGLQVPSSPYLLPGLVFSDFL"
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match 6.3%; Score 92; DB 6; Length 861;
Best Local Similarity 100.0%; Pred. No. 1.2e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 762 GGGCTGATGAGCGCTGCGGAGTGTCTGCGCAAGTTCGCGAGTGAAGCCAACTTCAAG 821
Db 463 GGGCTGATGAGCGCTGCGGAGTGTCTGCGCAAGTTCGCGAGTGAAGCCAACTTCAAG 522
Qy 822 ATCGAGACCTGCGCTTCCGCCCAACTACAT 853
Db 523 ATCGAGACCTGCGCTTCCGCCCAACTACAT 554

```

```

RESULT 6
MMU76208 861 bp DNA linear ROD 05-FEB-1997
LOCUS MMU76208
DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION U76208
VERSION U76208.1 GI:1815654
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 861)
Sommer, L., Ma, Q. and Anderson, D.J.
neurogenin, a novel family of atonal-related bHLH transcription
factors, are putative mammalian neuronal determination genes that
reveal progenitor cell heterogeneity in the developing CNS and PNS
Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
JOURNAL
MEDLINE
PUBMED
9000438
9715365
REFERENCE
2 (bases 1 to 861)
Sommer, L., Ma, Q. and Anderson, D.J.
Direct Submission
Submitted (24-OCT-1996) Biology 216-76, California Institute of
Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
USA
On Feb 5, 1997 this sequence version replaced gi:1666911.
COMMENT
FEATURES
source
Location/Qualifiers
1..861
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
160..804
/feature="ngn3"
160..804
/gene="ngn3"
/function="bHLH transcription factor"
/note="NGN3"
/codon_start=1

```

/product="neurogenin 3"  
/protein\_id="AAC53029.1"  
/db\_xref="GI:1666912"  
/translation="MAPHPDLALTIQVSEPTQPPGASDHEVLASNSSTPPTLIP  
DCSEAEVGDGCRGSTRKLRARGRGRNPKSELALSKORRRRRKANDRENNHNNLSA  
LDALRGVLPFPDPAKLTKEITLRFAHNITMALTOTLRADHSFGPEPPVCGEIGS  
PGGSGNDGWSIYSVQSQGNISPTASLEFPQLQVPSSPSYLPGALVFSDFL"

BASE COUNT 182 a 274 c 250 g 155 t

ORIGIN

Query Match 6.3%; Score 92; DB 10; Length 861;  
Best Local Similarity 100.0%; Pred. No. 1.2e-42;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGATGCGCTGCGCGGTGCTGCGCCACCTTCGCGATGAGCCCAACTTACAAG 821  
|||||  
DB 463 GCGCTGATGCGCTGCGCGGTGCTGCGCCACCTTCGCGATGAGCCCAACTTACAAG 522  
|||||

QY 822 ATCGAGACCCCTGCGCTTCCGCCCAACTACAT 853  
|||||  
DB 523 ATCGAGACCCCTGCGCTTCCGCCCAACTACAT 554  
|||||

RESULT 7  
AF364300 1861 bp DNA linear ROD 03-MAY-2001  
LOCUS Mus musculus neurogenin 3 gene, complete cds.  
DEFINITION  
AF364300  
ACCESSION  
AF364300.1 GI:13937128  
VERSION  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS Lee J., Smith S., Wacada H., Lin J., Schell D., Wang J., Mirmira R.  
and German M.  
TITLE Regulation of the pancreatic pro-endocrine gene neurogenin3  
JOURNAL Diabetes (2001) In press  
REFERENCE 2 (bases 1 to 1861)  
AUTHORS Schwitzgebel V. and German M.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAR-2001) Hormone Research Institute, University of  
California San Francisco, 513 Parnassus Ave, HSW1090, San  
Francisco, CA 94145-0534, USA  
FEATURES  
source location/Qualifiers  
1..1861  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
<1093..>1737  
/product="neurogenin 3"  
1093..1737  
/codon\_start=1  
/product="neurogenin 3"  
/protein\_id="AAK50058.1"  
/db\_xref="GI:13937129"  
/translation="MAPHPDLALTIQVSEPTQPPGASDHEVLASNSSTPPTLIP  
DCSEAEVGDGCRGSTRKLRARGRGRNPKSELALSKORRRRRKANDRENNHNNLSA  
LDALRGVLPFPDPAKLTKEITLRFAHNITMALTOTLRADHSFGPEPPVCGEIGS  
PGGSGNDGWSIYSVQSQGNISPTASLEFPQLQVPSSPSYLPGALVFSDFL"

BASE COUNT 397 a 560 c 537 g 367 t

ORIGIN

Query Match 6.3%; Score 92; DB 10; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 1.2e-42;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGATGCGCTGCGCGGTGCTGCGCCACCTTCGCGATGAGCCCAACTTACAAG 821  
|||||  
DB 1396 GCGCTGATGCGCTGCGCGGTGCTGCGCCACCTTCGCGATGAGCCCAACTTACAAG 1455  
|||||

QY 822 ATCGAGACCCCTGCGCTTCCGCCCAACTACAT 853  
|||||

DB 1456 ATCGAGACCCCTGCGCTTCCGCCCAACTACAT 1487  
|||||

RESULT 8  
MMA4TH4B 5567 bp DNA linear ROD 25-OCT-2000  
LOCUS MMA4TH4B  
DEFINITION M.musculus MMA4B gene.  
ACCESSION Y09167  
VERSION Y09167.2 GI:11065737  
KEYWORDS bHLH; mammalian atonal homolog 4B; MMA4B; neurogenin 3.  
SOURCE Mus musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS Cau E., Gradwohl G., Fode C. and Guillemot F.  
TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron  
JOURNAL Development 124 (8), 1611-1621 (1997)  
MEDLINE 97261963  
PUBMED 9108377

REFERENCE  
AUTHORS Jacquemin P., Durviaux S.M., Jensen J., Godfraind C., Gradwohl G.,  
Guillemot F., Madsen O.D., Carmeliet P., Dewerchin M., Collen D.,  
Rousseau G.G. and Lemaigre F.P.  
TITLE Transcription factor hepatocyte nuclear factor 6 regulates  
pancreatic endocrine cell differentiation and controls expression  
of the proendocrine gene ngn3  
JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)  
MEDLINE 20285449  
PUBMED 10825208

REFERENCE  
AUTHORS Gradwohl G.J.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC,  
CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg,  
F-67404 ILLKIRCH cedex, FRANCE  
REFERENCE 4 (bases 1 to 5567)  
AUTHORS Lemaigre F.P.  
TITLE Direct Submission  
JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic  
Research Unit, Louvain University Medical School, Avenue Hippocrate  
75, box 7529, Brussels 1200, BELGIUM  
COMMENT On Oct 31, 2000 this sequence version replaced gi:166087.  
FEATURES  
source location/Qualifiers  
1..5567  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="SV129"  
/db\_xref="taxon:10090"  
/clone="lambda GEM12 clone 19C"  
/cell\_line="D3"  
/cell\_type="embryonic stem cells"  
/clone\_lib="Genomic Library No.62, MboI partial digests  
cloned in BamHI"  
/dev\_stage="embryos"  
958  
/bound\_moiety="HNF6"  
3692  
/bound\_moiety="HNF6"  
4145..5567  
/gene="MMA4B"  
4145..4151  
/gene="MMA4B"  
/note="putative"  
4923..5567  
/gene="MMA4B"  
/codon\_start=1  
/product="neurogenin 3"  
/protein\_id="CAA70366.1"  
/db\_xref="GI:166088"

```

/db xref="SWISS-PROT:P70661"
/translation="MAPHPLDALTIQVSPETQPPGASDHEVLNSNSTPSPPTLIPR
DCEBAVGGCGCTGRKRRLRGRGRNRPSEILSKORSRGRKRRKRNHNHNSA
LDALRGVLPFPDDAKLTITETILRAHYHMYALTOTLRADPSFKGPPVPVCGILGS
PFGSGNDMGSIYSFVSQAGNISPTASLEFPGLQVPPSPITLLGALVSPFL"
BASE COUNT      1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match      6.3%; Score 92; DB 10; Length 5567;
Best Local Similarity 100.0%; Pred. No. 1,1e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGATGCGCTGCGCGGTCTCTCCACCTTCCCGATGAGCCCAACTTACAAG 821
Db 5226 GCGCTGATGCGCTGCGCGGTCTCTCCACCTTCCCGATGAGCCCAACTTACAAG 5285

Qy 822 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 853
Db 5286 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 5317

RESULT 9
AC109783/c      138070 bp      DNA      linear      HTG 07-FEB-2002
LOCUS
DEFINITION
Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.
AC109783
VERSION
AC109783.1 GI:18581594
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 138070)
McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zutavern,T.
Mouse Genomic Sequence
Unpublished
2 (bases 1 to 138070)
McCombie,W.R.
Direct Submission
Submitted (07-FEB-2002) Lite Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
----- Genome Center -----
Center: Lite Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mccombie@cshl.org
----- Project Information -----
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Insert size: 141616; sum-of-contrigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 17785: contrig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contrig of 12482 bp in length
* 30357 30444: gap of unknown length

```

```

* 30445 42306: contrig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contrig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64355: contrig of 10669 bp in length
* 64356 64443: gap of unknown length
* 64444 74016: contrig of 9573 bp in length
* 74017 74104: gap of unknown length
* 74105 83366: contrig of 9262 bp in length
* 83367 83455: gap of unknown length
* 83455 92355: contrig of 8901 bp in length
* 92356 92443: gap of unknown length
* 92444 100821: contrig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contrig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contrig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contrig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contrig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contrig of 4533 bp in length
* 128241 128328: gap of unknown length
* 128329 132682: contrig of 4354 bp in length
* 132683 132770: gap of unknown length
* 132771 136341: contrig of 3571 bp in length
* 136342 136429: gap of unknown length
* 136430 138070: contrig of 1641 bp in length.
location/Qualifiers
source
1. 138070
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-121F10"

BASE COUNT      35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Query Match      6.3%; Score 92; DB 2; Length 138070;
Best Local Similarity 100.0%; Pred. No. 9.7e-43;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGATGCGCTGCGCGGTCTCTGCCCACTTCCCGATGAGCCCAACTTACAAG 821
Db 110917 GCGCTGATGCGCTGCGCGGTCTCTGCCCACTTCCCGATGAGCCCAACTTACAAG 110858

Qy 822 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 853
Db 110857 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 110826

RESULT 10
AC127417      215050 bp      DNA      linear      HTG 19-OCT-2002
LOCUS
DEFINITION
Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC127417
AC127417.2 GI:24137619
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FILLTOP.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 215050)

```

ORIGIN

JOURNAL	REMARK COMMENT
17108	Direct submission Submitted (23-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA  NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>  Contact: MGC help desk Email: <a href="mailto:cgaabps-remail.nih.gov">cgaabps-remail.nih.gov</a> Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: <a href="http://www.nisc.nih.gov/">http://www.nisc.nih.gov/</a> Contact: <a href="mailto:nisc.mgc@nhgri.nih.gov">nisc.mgc@nhgri.nih.gov</a> Akter, N., Ayele, K., Beckerson-Sternberg, S.M., Benjamin, B., Blakeley, R.W., Bouffard, G.G., Breem, K.C., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, U., Hashiguchi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, J., Maduro, Q., Masello, C., Maekeri, B., Meitner, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Scantiripoli, S., Thomas, P.D., Touchman, J.W.



Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/URL at: <http://image.llnl.gov>  
Series: IRAC Plate: 78 Row: K Column: 16.  
Location/Qualifiers

## FEATURES

source

1. 2370  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone\_id="MGC:46562 IMAGE:5247719"  
/cissue\_type="Brain, fetal, whole pooled"  
/clone\_lib="NIH MGC\_121"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 2370  
/gene="NEUROG2"  
/note="synonyms: Atoh4, Math4a, ngn-2, MGC46562"  
/db\_xref="LOCUSTID:63973"  
/db\_xref="MIM:606624"  
328. 1146  
/codon\_start=1  
/product="neurogenin 2"  
/protein\_id="AAH36847.1"  
/db\_xref="GI:22477417"  
/db\_xref="LOCUSTID:63973"  
/translation="MFVKSETLFLKEEDVVLIGSASPALAATPLSSADEEBEE  
PASGGARRQGAEGAGARGVAGAECCRPRLGLVHDCRRPRRAAARAKT  
AETVQRICKTRRLKANNRRNMHNALDLREVLPTPEPAKTKLETLRFANY  
IWALETTLADHCGGGGLPGALFSEAVLISPGASALSSGSDSPSPASTWSTN  
SPAPSSSVSNSTSPYSCITLSPASGSDMDWQPPPKHRYAPHLPIARDCI"

## CDS

BASE COUNT 580 a 654 c 616 g 520 t  
ORIGIN

Query Match 2.4%; Score 35; DB 9; Length 2370;  
Best Local Similarity 100.0%; Pred. No. 2.1e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853  
DB 775 AAGATCGAGACCTCGGCTTGGCCCACTACAT 809

RESULT 12  
AF303002 6123 bp DNA linear PRI 13-NOV-2001  
LOCUS Homo sapiens neurogenin 2 gene, partial cds.  
DEFINITION AF303002  
ACCESSION AF303002  
VERSION AF303002.1 GI:11875763  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 6123)  
AUTHORS Simmons, A.D., Horton, S., Abney, A.L. and Johnson, J.E.  
TITLE Neurogenin2 expression in ventral and dorsal spinal neural tube  
progenitor cells is regulated by distinct enhancers  
JOURNAL Dev. Biol. 229 (2), 327-339 (2001)  
MEDLINE 11203697  
PUBMED 11203697

REFERENCE 2 (bases 1 to 6123)  
AUTHORS Simmons, A.D., Horton, S., Abney, A.L. and Johnson, J.E.  
TITLE Direct Submission  
JOURNAL Submitted (06-SEP-2000) Center for Basic Neuroscience - N44.146, UT  
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX  
75390-9111, USA

FEATURES  
source  
1. 6123  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

mRNA  
CDS  
1. .754  
/product="neurogenin 2"  
1. .754  
/note="NM2"  
/codon\_start=2  
/product="neurogenin 2"  
/protein\_id="AA040770.1"  
/db\_xref="GI:11875764"  
/translation="ASPALAATPLSSADEEBEEPASGGARRQGAEGARGG  
VAAGGCRPARLGLVHDCRRPRRAAARAKTETVQRICKTRRLKANNRRN  
MHNALDLREVLPTPEPAKTKLETLRFANYIWALETTLADHCGGGGLP  
GALFSEAVLISPGASALSSGSDSPSPASTWSTN  
SPAPSSSVSNSTSPYSCITLSPASGSDMDWQPPPKHRYAPHLPIARDCI"

BASE COUNT 1484 a 1536 c 1507 g 1596 t  
ORIGIN

Query Match 2.4%; Score 35; DB 9; Length 6123;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853  
DB 383 AAGATCGAGACCTCGGCTTGGCCCACTACAT 417

RESULT 13  
AC079846 3/c  
WPCOMMENT  
Sequence split into 4 fragments LOCUS AC079846 Accession AC079846  
Fragment Name Begin End  
AC079846\_1 1 110000  
AC079846\_2 100001 210000  
AC079846\_3 200001 310000  
AC079846\_4 300001 391531

Continuation (4 of 4) of AC079846 from base 300001 (AC079846 Homo sapiens chromosome 3 c  
Query Match 2.4%; Score 35; DB 2; Length 91531;  
Best Local Similarity 100.0%; Pred. No. 1.8e-08;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853  
DB 86995 AAGATCGAGACCTCGGCTTGGCCCACTACAT 86961

RESULT 14  
AC023886/c  
LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002  
DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.  
ACCESSION AC023886  
VERSION AC023886.7 GI:19482381  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 179697)  
AUTHORS Sulston, J.E. and Waterston, R.  
TITLE Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 9847074  
PUBMED 9847074

REFERENCE 2 (bases 1 to 179697)  
AUTHORS Radionenko, M. and Abbott, A.  
TITLE The sequence of Homo sapiens BAC clone RP11-402J6  
JOURNAL Unpublished (2001)  
REFERENCE 3 (bases 1 to 179697)  
AUTHORS Waterston, R.H.

REFERENCE 4 (bases 1 to 179697)  
JOURNAL Direct Submission  
Submitted (18-FEB-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA



AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 5 (bases 1 to 179697)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Mar 15, 2002 this sequence version replaced gi:17352441.

----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.wustl.edu](mailto:sapiens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH040206  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is AC004049, 2000 bp overlap, the clone sequenced to the right is RP11-14886. Actual end of this clone is at base position 179697 of RP11-40206.  
Location/Qualifiers

## FEATURES

## source

1. 179697  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-40206"  
/clone\_1ib="RPCT-11"  
1. 749  
/rpt\_family="L1"  
repeat\_region  
734. 3067  
/rpt\_family="L1"  
repeat\_region  
3104. 3444  
/rpt\_family="MALR"  
repeat\_region  
3447. 4289  
/rpt\_family="L1"  
repeat\_region  
4290. 4369  
/rpt\_family="TA)n"  
repeat\_region  
4370. 4797

/rpt\_family="L1"  
repeat\_region  
5022. 5194  
/rpt\_family="L2"  
repeat\_region  
5214. 5628  
/rpt\_family="L1"  
repeat\_region  
5710. 5729  
/rpt\_family="TA)n"  
repeat\_region  
5749. 5863  
/rpt\_family="TA)n"  
repeat\_region  
5883. 6263  
/rpt\_family="ERV.L"  
repeat\_region  
9901. 10151  
/rpt\_family="ERV.L"  
repeat\_region  
10152. 10310  
/rpt\_family="L1"  
repeat\_region  
10311. 10400  
/rpt\_family="Mariner"  
repeat\_region  
10401. 10816  
/rpt\_family="L1"  
repeat\_region  
10967. 12362  
/rpt\_family="L1"  
repeat\_region  
12366. 12653  
/rpt\_family="Alu"  
repeat\_region  
12695. 13940  
/rpt\_family="L1"  
repeat\_region  
13965. 14272  
/rpt\_family="Alu"  
repeat\_region  
14274. 14795  
/rpt\_family="ERV.L"  
repeat\_region  
14806. 15192  
/rpt\_family="L1"  
repeat\_region  
21504. 22112  
/rpt\_family="L1"  
repeat\_region  
22131. 23166  
/rpt\_family="L1"  
repeat\_region  
23246. 23386  
/rpt\_family="MALR"  
repeat\_region  
23511. 23654  
/rpt\_family="MIR"  
repeat\_region  
23655. 25486  
/rpt\_family="L1"  
repeat\_region  
25487. 25636  
/rpt\_family="MIR"  
repeat\_region  
25814. 26124  
/rpt\_family="Alu"  
repeat\_region  
26444. 26592  
/rpt\_family="MIR"  
repeat\_region  
27423. 27713  
/rpt\_family="Alu"  
repeat\_region  
27716. 27739  
/rpt\_family="AT\_rich"  
repeat\_region  
27773. 27795  
/rpt\_family="AT\_rich"  
repeat\_region  
28020. 28429  
/rpt\_family="ERV.L"  
repeat\_region  
29646. 29668  
/rpt\_family="AT\_rich"  
repeat\_region  
29727. 29871  
/rpt\_family="L1"  
repeat\_region  
29875. 30114  
/rpt\_family="Alu"  
repeat\_region  
30134. 30267  
/rpt\_family="L1"  
repeat\_region  
30268. 30549  
/rpt\_family="L1"  
repeat\_region  
30550. 30883  
/rpt\_family="MALR"  
repeat\_region  
32461. 32791  
/rpt\_family="MALR"  
repeat\_region  
32792. 33487  
/rpt\_family="L1"  
repeat\_region  
33488. 33847  
/rpt\_family="MALR"

```

repeat_region      33848..34354
                    /rpt_family="L1"
repeat_region      34358..34379
                    /rpt_family="(CATATA)n"
repeat_region      34381..34502
                    /rpt_family="Alu"
repeat_region      34503..34556
                    /rpt_family="L1"
repeat_region      34557..34859
                    /rpt_family="Alu"
repeat_region      34860..35041
                    /rpt_family="L1"
repeat_region      35219..35676
                    /rpt_family="L2"
repeat_region      35974..36042
                    /rpt_family="L2"
repeat_region      36833..37144
                    /rpt_family="Alu"
misc_feature        36833..36848
                    /note="match to EST AY537116 (NID:g4451251) t01sh09.x1"
repeat_region      37147..37410
                    /rpt_family="Alu"
repeat_region      38154..38424
                    /rpt_family="Alu"
repeat_region      39941..40230
                    /rpt_family="Alu"
misc_feature        39941..39956
                    /note="match to EST AA75240 (NID:g2834574) ad18a05.s1"

```

```

Query Match      2.4%; Score 35; DB 9; Length 179697;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      819 AAGATGAGACCTGCGCTTGGCCGACACTACAT 853
Db      55815 AAGATGAGACCTGCGCTTGGCCGACACTACAT 55781

```

```

RESULT 15
LOCUS      HSA133776              1330 bp      DNA              linear      PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION  AJ133776
VERSION     AJ133776.1 GI:5123782
KEYWORDS   bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE   1
AUTHORS     Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
TITLE       The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterpart
            Unpublished
            2 (bases 1 to 1330)
JOURNAL     Direct Submission
AUTHORS     Ravassard,P.
TITLE       Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMR 9923, Hopital
            de la Pitie Salpêtrière, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE
FEATURES
source      1..1330
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            1..1330
            /gene="ngn3"
            join(<1..157,321..>1330)
            /gene="ngn3"
            1..157
            /gene="ngn3"
            /number=1

```

```

intron          158..320
                /gene="ngn3"
                /number=1
exon            321..1330
                /gene="ngn3"
                /number=2
                322..366
                /gene="ngn3"
                /function="early neurogenesis"
                /codon_start=1
                /product="neurogenin 3"
                /protein_id="CAB45384.1"
                /db_xref="GI:5123783"
                /translation="MTQPSGAPTVQVTRTERSPPRASEDEVTCPTISAPSPTRTPG
                NCAEAEEGCGAPRKLARRGSRSPKSEIALSKORSRKANDRRRMHDLNSA
                LDALRGVPTFPDDAKLTIETLRFAHYIWALETIADHSIYALEPPAPHGCELG
                SPGGRPGDMSLSPVSGAGSLSPASIEERGLIGATSSACLSBGLAFSDFL"
BASE COUNT      230 a      459 c      413 g      228 t
ORIGIN

```

```

Query Match      2.2%; Score 32; DB 9; Length 1330;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      672 AAGACGAGTTGGCACTGACGACGACG 703
Db      535 AAGACGAGTTGGCACTGACGACGACG 566

```

```

Search completed: January 29, 2004, 21:29:57
Job time : 5314 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:49:27 ; Search time 117 Seconds  
(without alignments)  
5507.858 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gtcagtagcagagagagcag.....agagtgcactcaccagtc 1460.

Scoring table: OLIGO\_NUC

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	6.3	804	US-08-932-411A-19	Sequence 19, Appl
2	23	1.6	738	US-08-722-570-13	Sequence 13, Appl
3	23	1.6	738	US-08-932-411A-13	Sequence 13, Appl
4	23	1.6	1333	US-08-910-973-21	Sequence 21, Appl
5	23	1.6	1333	US-08-910-973-21	Sequence 21, Appl
6	23	1.6	1385	US-08-932-411A-17	Sequence 17, Appl
7	21	1.4	310	US-08-552-142A-12	Sequence 12, Appl
8	21	1.4	1268	US-08-910-973-12	Sequence 12, Appl
9	21	1.4	1352	US-08-552-142A-10	Sequence 10, Appl
10	21	1.4	1352	US-08-552-142A-10	Sequence 10, Appl
11	21	1.4	1535	US-08-910-973-10	Sequence 10, Appl
12	21	1.4	1535	US-08-910-973-10	Sequence 10, Appl
13	21	1.4	1550	US-09-234-332-3	Sequence 3, Appl
14	19	1.3	50	US-08-358-627F-4	Sequence 4, Appl
15	19	1.3	50	US-08-793-044-11	Sequence 11, Appl
16	19	1.3	50	US-08-465-712C-4	Sequence 4, Appl
17	19	1.3	50	US-08-465-712C-4	Sequence 4, Appl
18	18	1.2	480	US-08-438-123-16	Sequence 16, Appl
19	18	1.2	1336	US-09-016-434-129	Sequence 129, App
20	18	1.2	1462	US-08-552-142A-16	Sequence 16, Appl
21	18	1.2	1951	US-08-910-973-16	Sequence 16, Appl
22	18	1.2	1951	US-08-910-973-16	Sequence 16, Appl
23	18	1.2	3446	US-09-499-227-16	Sequence 16, Appl
24	18	1.2	3446	US-09-499-227-16	Sequence 16, Appl
25	18	1.2	4066	US-09-367-750-1	Sequence 1, Appl
26	18	1.2	4797	US-09-419-568F-25	Sequence 25, Appl
27	18	1.2	4797	US-09-354-243B-25	Sequence 25, Appl
27	18	1.2	31728	US-09-453-702B-64	Sequence 64, Appl

28	17	1.2	24	3	US-08-358-627F-6	Sequence 6, Appl
29	17	1.2	24	4	US-08-465-712C-6	Sequence 6, Appl
30	17	1.2	24	4	US-09-552-733-6	Sequence 6, Appl
31	17	1.2	24	4	US-09-687-731-2	Sequence 2, Appl
32	17	1.2	24	4	US-09-687-731-2	Sequence 2, Appl
33	17	1.2	24	4	US-09-687-731-8	Sequence 8, Appl
34	17	1.2	24	4	US-09-687-731-10	Sequence 10, Appl
35	17	1.2	510	4	US-08-252-991A-9622	Sequence 9622, Ap
36	17	1.2	685	4	US-08-751-782-5	Sequence 5, Appl
37	17	1.2	685	2	US-08-925-171-5	Sequence 9635, Ap
38	17	1.2	846	4	US-09-252-991A-9635	Sequence 14810, A
39	17	1.2	954	4	US-09-252-991A-14810	Sequence 1994, Ap
40	17	1.2	1062	4	US-09-252-991A-9532	Sequence 9532, Ap
41	17	1.2	1161	1	US-08-086-439C-1	Sequence 1, Appl
42	17	1.2	1161	1	US-08-434-877-1	Sequence 1, Appl
43	17	1.2	1238	4	US-09-183-861-75	Sequence 75, Appl
44	17	1.2	1238	4	US-09-022-765-75	Sequence 75, Appl
45	17	1.2	1238	4	US-09-551-974A-75	Sequence 75, Appl

#### ALIGNMENTS

RESULT 1  
US-08-932-411A-19  
Sequence 19, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qinfu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Flehr Hobbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RPT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 804 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 160..801  
US-08-932-411A-19

Query Match 6.3%; Score 92; DB 4; Length 804;  
Best Local Similarity 100.0%; Pred. No. 9.9e-37;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GCGGTGATGGCTGGGGGTCTCTGCCACCTTCCGGATGAGCCCAACTTCAAG 821  
DB 463 GCGGTGATGGCTGGGGGTCTCTGCCACCTTCCGGATGAGCCCAACTTCAAG 522  
OY 822 ATCGAGACCTGCGCTTGGCCCAACTACAT 853  
DB 523 ATCGAGACCTGCGCTTGGCCCAACTACAT 554

## RESULT 2

US-08-722-570-13  
Sequence 13, Application US/08722570  
Patent No. 6555337  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Olufa  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/722,570  
FILING DATE: 27-SEP-1996  
CLASSIFICATION: 5365  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 738 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
US-08-722-570-13

Query Match 1.6%; Score 23; DB 4; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754  
DB 304 GAGCGCAACCGCATGCACACCT 326

## RESULT 3

US-08-932-411A-13  
Sequence 13, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Olufa  
TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/772,009

FILING DATE: 19-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/722,570

FILING DATE: 19-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 738 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..732

US-08-932-411A-13

Query Match 1.6%; Score 23; DB 4; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754  
DB 304 GAGCGCAACCGCATGCACACCT 326

## RESULT 4

US-08-910-973-21  
Sequence 21, Application US/08910973  
Patent No. 5795723  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910.973  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239.238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FPCR-1-10958  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1333 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
IMMEDIATE SOURCE:  
CLONE: neuroD3  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 101..835  
US-08-910-973-21

Query Match 1.6%; Score 23; DB 1; Length 1333;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754  
|||||  
Db 404 GAGCGCAACCGCATGCACAACT 426

RESULT 5  
US-09-499-227-21  
Sequence 21, Application US/09499227  
Patent No. 644463  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/499,227  
FILING DATE: 05-August-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-May-1994  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-May-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910.973  
FILING DATE: 07-August-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FPCR-1-12742  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1333 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Mus musculus  
IMMEDIATE SOURCE:  
CLONE: neuroD3  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 101..835  
US-09-499-227-21

Query Match 1.6%; Score 23; DB 4; Length 1333;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754  
|||||  
Db 404 GAGCGCAACCGCATGCACAACT 426

RESULT 6  
US-08-932-411A-17  
Sequence 17, Application US/08932411A  
Patent No. 6566496  
GENERAL INFORMATION:  
APPLICANT: Anderson, David J.  
APPLICANT: Ma, Qifu  
TITLE OF INVENTION: NEUROGENIN  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP  
STREET: Four Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/932,411A  
FILING DATE: 15-SEP-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/772,009  
FILING DATE: 19-DEC-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/722,570  
FILING DATE: 19-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1385 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 382..1170  
US-08-932-411A-17

Query Match 1.6%; Score 23; DB 4; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 0.072;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754  
DB 742 GAGCGCAACCGCATGCACACCT 764

RESULT 7  
US-08-552-142A-12  
Sequence 12, Application US/08552142A  
Patent No. 5695995  
GENERAL INFORMATION:  
APPLICANT: Weintraub, Harold M.  
APPLICANT: Lee, Jacqueline E.  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Hollenberg, Stanley M.  
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/552,142A  
FILING DATE: 02-NOV-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas P.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FHCR-1-8933  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100  
TELEFAX: 206-225-0709  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 310 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 20A1  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..310  
US-08-552-142A-12

Query Match 1.4%; Score 21; DB 1; Length 310;  
Best Local Similarity 100.0%; Pred. No. 0.72;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752  
DB 128 GAGCGCAACCGCATGCACAC 148

RESULT 8  
US-08-910-973-12  
Sequence 12, Application US/08910973  
Patent No. 5795723  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,973  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheinets, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FHCR-1-10958  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1268 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 20A1 (neuroD3)  
FEATURE:

NAME/KEY: CDS  
LOCATION: 55..768  
US-08-910-973-12

Query Match 1.4%; Score 21; DB 1; Length 1268;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752  
|||||  
DB 355 GAGCGCAACCGCATGCACAC 375

RESULT 9  
US-09-499-227-12

Sequence 12, Application US/09499227  
Patent No. 6444463

GENERAL INFORMATION:

APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.

TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen O'Connor Johnson KindnessPLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/499,227

FILING DATE: 05-August-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-May-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-May-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/17532

FILING DATE: 30-October-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/910,973

FILING DATE: 07-August-1997

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: FHCR-1-12742

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100; 206-224-0735 (direct)

TELEFAX: 206-225-0779

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1268 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: 20A1 (neuroD3)

FEATURE:

NAME/KEY: CDS

LOCATION: 55..768

US-09-499-227-12

Query Match

1.4%; Score 21; DB 4; Length 1268;

Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752  
|||||  
DB 355 GAGCGCAACCGCATGCACAC 375

RESULT 10  
US-08-552-142A-10

Sequence 10, Application US/08552142A  
Patent No. 5695995

GENERAL INFORMATION:

APPLICANT: Weintraub, Harold M.

APPLICANT: Lee, Jacqueline E.

APPLICANT: Tapscott, Stephen J.

APPLICANT: Hollenberg, Stanley M.

TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen O'Connor Johnson KindnessPLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/552,142A

FILING DATE: 02-NOV-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-MAY-1995

ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.

REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: FHCR-1-8933

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100

TELEFAX: 206-225-0709

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1352 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: 14B1

FEATURE:

NAME/KEY: CDS

LOCATION: 55..1194

US-08-552-142A-10

Query Match 1.4%; Score 21; DB 1; Length 1352;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 729 CGGAGCGCAACCGCATGCAC 749  
|||||  
DB 439 CGGAGCGCAACCGCATGCAC 459

Query Match 1.4%; Score 21; DB 1; Length 1352;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11  
US-08-910-973-10  
Sequence 10, Application US/08910973  
Patent No. 5795723  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.  
APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson KindnessAPLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910,973  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FHCR-1-10958  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1535 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 14B1 (neuroD2)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 55..1194  
US-08-910-973-10

Query Match 1.4%; Score 21; DB 1; Length 1535;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CCGGAGCGCAACCGCATGCAC 749  
DB 439 CCGGAGCGCAACCGCATGCAC 459

RESULT 12  
US-09-499-227-10  
Sequence 10, Application US/09499227  
Patent No. 6444463  
GENERAL INFORMATION:  
APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.  
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen O'Connor Johnson KindnessAPLLC  
STREET: 1420 Fifth Avenue, Suite 2800  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/499,227  
FILING DATE: 05-August-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/239,238  
FILING DATE: 06-MAY-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US95/05741  
FILING DATE: 08-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/17532  
FILING DATE: 30-October-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/910,973  
FILING DATE: 07-August-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Sheiness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: FHCR-1-12742  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-682-8100; 206-224-0735 (direct)  
TELEFAX: 206-225-0779  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1535 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: 14B1 (neuroD2)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 55..1194  
US-09-499-227-10

Query Match 1.4%; Score 21; DB 4; Length 1535;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CCGGAGCGCAACCGCATGCAC 749  
DB 439 CCGGAGCGCAACCGCATGCAC 459

RESULT 13  
US-09-234-332-3  
Sequence 3, Application US/09234332A  
Patent No. 6087168  
GENERAL INFORMATION:  
APPLICANT: Cedars-Sinai Medical Center  
APPLICANT: Michael F. Levesque, M.D.  
APPLICANT: Thomas Neuman, Ph.D.  
TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO  
TITLE OF INVENTION: NEURONS; TRANSDIFFERENTIATION OF EPIDERMAL CELLS  
FILE REFERENCE: P07 41494



CURRENT APPLICATION NUMBER: US/09/234,332A  
CURRENT FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 3  
LENGTH: 1550  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: gene  
LOCATION: (0)...(0)  
OTHER INFORMATION: Neurogenic basic-helix-loop-helix protein (Neuro  
OTHER INFORMATION: D2) gene Genbank Accession U58681  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1219)...(1226)  
OTHER INFORMATION: n at 1219 and 1226; n = A, T, G, or C  
US-09-234-332-3

Query Match 1.4%; Score 21; DB 3; Length 1550;  
Best Local Similarity 100.0%; Pred. No. 0.74;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 729 CGGAGCGCACCGCATGCAC 749  
Db 442 CGGAGCGCACCGCATGCAC 462

RESULT 14  
US-08-358-627F-4/C  
Sequence 4, Application US/08358627F  
Patent No. 6177242  
GENERAL INFORMATION:  
APPLICANT: Changeux, Jean-Pierre  
APPLICANT: Picciotto, Marina  
TITLE OF INVENTION: Genomic DNA Fragments Containing  
TITLE OF INVENTION: Regulatory and Coding Sequences for the R2-Subunit of the  
TITLE OF INVENTION: Neuronal Nicotinic Acetylcholine Receptor and Transgenic  
TITLE OF INVENTION: Animals Made Using These Fragments or Mutated Fragments  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Flumegan, Henderson, Farabow, Garrett &  
ADDRESSEE: Dunner, L.L.P.  
STREET: 1300 I Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/358,627F  
FILING DATE: 14-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0135-00000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-358-627F-4

Query Match 1.3%; Score 19; DB 3; Length 50;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCAGGTAGCGAGGAGCA 19  
Db 33 GCAGGTAGCGAGGAGCA 15

RESULT 15  
US-08-793-044-11/C  
Sequence 11, Application US/08793044  
Patent No. 6235497  
GENERAL INFORMATION:  
APPLICANT: Benjamin, Stephane  
APPLICANT: Berrard, Sylvie  
APPLICANT: Cerrini, Riccardo  
APPLICANT: Maillet, Jacques  
TITLE OF INVENTION: NOVEL VESICULAR ACETYLCHOLINE CARRIER  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, Mailstop 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,044  
FILING DATE:  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/01073  
FILING DATE: 10-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/10044  
FILING DATE: 16-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Savitzky Esq., Martin F.  
REGISTRATION NUMBER: 29,699  
REFERENCE/DOCKET NUMBER: ST94066-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3816  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
US-08-793-044-11

Query Match 1.3%; Score 19; DB 3; Length 50;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GCAGGTAGCGAGGAGCA 19  
Db 33 GCAGGTAGCGAGGAGCA 15

Search completed: January 29, 2004, 22:20:35  
Job time : 119 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 19:39:06 ; Search time 426 Seconds

(without alignments)  
9251.595 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460  
1 gcagtgacgagagagagcag.....agagtgaactaatccagtgt 1460

Sequence: OLIGO NUC

Scoring table: Gapop 60.0, Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

N\_Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1409	96.5	1491	19	AAV42512
2	92	6.3	804	19	AAV27050
3	92	6.3	804	21	AAZ51981
4	92	6.3	861	22	AAF27266
5	92	6.3	861	25	ABV75970
6	92	6.3	1860	24	AAD46872
7	92	6.3	1861	21	AAC61090
8	92	6.3	5567	22	AAF27254

9	36	2.5	65	24	ABN1392
10	35	2.4	65	24	ABN57521
11	35	2.4	6123	24	AAD46890
12	32	2.2	5340	21	AAC61089
13	32	2.2	5340	24	AAD46871
14	29	2.0	428	22	AAS33797
15	26	1.8	26	24	ABT03700
16	25	1.7	25	19	AAV42515
17	23	1.6	738	19	AAV27046
18	23	1.6	738	21	AAZ51977
19	23	1.6	790	22	AAF27264
20	23	1.6	1332	19	AAV42938
21	23	1.6	1332	25	AB556396
22	23	1.6	1333	18	AAV74894
23	23	1.6	1385	21	AAZ51980
24	23	1.6	1385	22	AAF27269
25	23	1.6	1385	22	AAF27255
26	23	1.6	1412	22	AAF27255
27	23	1.6	1412	22	AAF27273
28	21	1.4	21	25	AAD47278
29	21	1.4	21	25	AAD47326
30	21	1.4	714	24	AAD46889
31	21	1.4	1268	18	AAV74891
32	21	1.4	1268	19	AAV42932
33	21	1.4	1268	25	AB556390
34	21	1.4	1535	18	AAV74890
35	21	1.4	1535	21	AAV42931
36	21	1.4	1535	25	AB556389
37	21	1.4	1550	21	AAAC2681
38	21	1.4	1665	24	AAV46888
39	21	1.4	2776	22	AAV4043
40	21	1.4	2776	22	AAV4045
41	21	1.4	2776	22	AAK68475
42	21	1.4	2776	22	AAK68475
43	20	1.4	352	24	AB569045
44	20	1.4	592	24	ABQ49522
45	20	1.4	592	24	ABQ49523

## ALIGNMENTS

RESULT 1	AAV42512	standard; cDNA; 1491 BP.
ID	AAV42512	
AC	AAV42512;	
DT	05-OCT-1998	(first entry)
DE	CDNA encoding a novel BHLH protein designated RELAX.	
XX	Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;	
KW	control; gene expression; transcriptional activator; targeting;	
KW	protein expression; central nervous system; CNS; treatment;	
KW	nervous system disorder; ss.	
OS	Rattus sp.	
XX		
FH	Key	Location/Qualifiers
FT	459...1103	/tag= a
FT	CDS	/product= RELAX
XX		
PN	W09827206-A2.	
XX		
PD	25-JUN-1998.	
XX		
PF	19-DEC-1997;	97WO-FR02368.
XX		
PR	19-DEC-1996;	96FR-0015651.
XX		
PA	(RHON ) RHONE-POULENC RORER SA.	

Rat spliced transc  
Mouse spliced tran  
Human neurogenin 2  
Human neurogenin 3  
Human neurogenin 3  
Human cDNA encodin  
Human Neurogenin-3  
PCR primer used to  
Mouse neurogenin-1  
Murine neurogenin-  
Chicken atonal hom  
DNA encoding murin  
Mouse BHLH family  
Mouse neurogenic d  
Mouse neurogenic d  
Murine neurogenin-  
Mouse neurogenin-2  
Mouse atonal homol  
Mouse atonal homol  
Human RT-PCR upstr  
Human RT-PCR forwa  
Human neurogenin 1  
Human neurogenic d  
DNA encoding human  
Human bHLH family  
Human neurogenic d  
DNA encoding human  
Human bHLH family  
Human NeuroD2 gene  
Human neurogenin 1  
Human reproductive  
Human reproductive  
Human immune/haema  
Novel murine polyn  
Oligonucleotide fo  
Oligonucleotide fo



FT /product= "Mouse neurogenin 3"

XX WO9813491-A2.

XX 02-APR-1998.

XX 24-SEP-1997; 97WO-US17048.

XX 17-SEP-1997; 97US-0932411.

XX 27-SEP-1996; 96US-0722570.

XX 12-NOV-1996; 96US-0030864.

XX 19-DEC-1996; 96US-0772009.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Ma Q, Sommer L;

XX WPI; 1998-230702/20.

XX P-PSDB; AAM54947.

XX Mouse neurogenin, useful in neurogenesis - and recombinant nucleic acids and proteins derived from rat and Xenopus

XX Disclosure; Fig 9, 106pp; English.

XX The Mouse neurogenin 3 is one of several neurogenin proteins discussed in the present invention. The neurogenin nucleic acids can be expressed in a host cell, transformed using an expression vector, to produce recombinant proteins. The proteins and the antibodies raised against the proteins are useful in the study of neurogenesis.

XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

SQ Query Match 6.3%; Score 92; DB 19; Length 804;

Best Local Similarity 100.0%; Pred. No. 9e-35;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGATGCGCTGCGGCTGCTGCGCACTTCCCGATGAGCCCAACTTACAAG 821

DB 463 GCGCTGATGCGCTGCGGCTGCTGCGCACTTCCCGATGAGCCCAACTTACAAG 522

QY 822 ATCGAGACCTTGCGCTTGCGCCCAACTACAT 853

DB 523 ATCGAGACCTTGCGCTTGCGCCCAACTACAT 554

RESULT 3

AAZ51981

ID AAZ51981 standard; DNA; 804 BP.

XX AAZ51981;

XX 04-JUN-2000 (first entry)

DE Murine neurogenin-3 (NGN3) nucleic acid sequence.

XX Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;

XX Phox2a protein; neuronal subtype-specific marker; growth factor;

XX neuronal differentiation; transplantation; neuronal dysfunction;

XX optical nerve damage; auditory nerve damage; neurodegenerative disorder;

XX neuroprotective; neurotropic; anticonvulsant; antiParkinsonian; vulnerary;

XX cerebroprotective; immunosuppressant; antiinfectious; ds.

XX Mus sp.

OS Location/Qualifiers

XX Key 160..804

XX CDS /\*tag= a

FT /product= "Murine neurogenin-3 protein"

XX WO200009676-A2.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18525.

XX 14-AUG-1998; 98US-0096630.

XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Lo L;

XX WPI; 2000-256250/22.

XX P-PSDB; AAY70570.

XX Inducing non-neuronal cells to differentiate into neurons and for non-neuronal cells to express a neuronal subtype-specific marker, comprising contacting the non-neuronal cells with a vector containing neurogenin nucleic acid -

XX Claim 1; Fig 1J; 76pp; English.

XX The patent discloses a method for inducing non-neuronal cells (NNC) to differentiate into neurons and for NNCs to express a neuronal subtype-specific marker. Transformed host cells are used as sources of neuronal and other growth factors, in culture for screening compounds that modulate neural differentiation or as sources of recombinantly produced neurogenins and Phox2a proteins for use in transplantation. The cells also have a variety of in vivo uses, e.g. for transplantation at sites of neuronal dysfunction e.g. patients with hearing or vision loss due to optical or auditory nerve damage, brain or spinal cord injuries, and neurodegenerative disorders e.g. Alzheimer's disease. The present sequence encodes murine neurogenin-3 (NGN-3), a transcription factor. CC NNCs differentiate into neurons through the recombinant expression of a transcription factor that induces a core program of neurogenesis. Forced expression of murine NGN3 can elicit expression of at least some neuronal phenotypic markers even in NNCs.

XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

SQ Query Match 6.3%; Score 92; DB 21; Length 804;

Best Local Similarity 100.0%; Pred. No. 9e-35;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGATGCGCTGCGGCTGCTGCGCACTTCCCGATGAGCCCAACTTACAAG 821

DB 463 GCGCTGATGCGCTGCGGCTGCTGCGCACTTCCCGATGAGCCCAACTTACAAG 522

QY 822 ATCGAGACCTTGCGCTTGCGCCCAACTACAT 853

DB 523 ATCGAGACCTTGCGCTTGCGCCCAACTACAT 554

RESULT 4

AAF27266

ID AAF27266 standard; cDNA; 861 BP.

XX AAF27266;

XX 24-APR-2001 (first entry)

DE Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;

XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;

XX cellular proliferation; cerebellar granule neuron; gene therapy;

XX mechanoreceptive cell growth; auditory; osteopathic; cyostatic;

XX transgenic animal; ss.

XX Mus musculus.

XX WO2000073764-A2.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.

XX 01-JUN-1999; 99US-0137060.  
 PR 19-JAN-2000; 2000US-0176993.  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA Zoghbi HY, Ballen H, Birmingham N, Hassan B, Ben-Arie N;  
 PI MPI; 2001-032190/04.  
 DR P-PSDB; AAB60359.  
 XX  
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation -  
 XX  
 PS Disclosure; Page -; 142pp; English.  
 XX  
 CC The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologues or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal,  
 CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from GenBank.  
 XX  
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;  
 QY  
 Query Match 6.3%; Score 92; DB 22; Length 861;  
 Best Local Similarity 100.0%; Pred. No. 9e-35;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 762 GGGCTGATGCGCTGCGGCTGCTCCGACCTTCCGGATGAGCGCAACTTACAAG 821  
 463 GGGCTGATGCGCTGCGGCTGCTCCGACCTTCCGGATGAGCGCAACTTACAAG 522  
 QY 822 ATCGAGACCTTGCGCTTCCGCCCAACTACAT 853  
 Db 523 ATCGAGACCTTGCGCTTCCGCCCAACTACAT 554  
 RESULT 5  
 ABV75970  
 ID ABV75970 strand; cDNA; 861 BP.  
 AC ABV75970;  
 XX  
 DT 11-FEB-2003 (first entry)  
 DE Mouse transcription factor neurogenin 3 cDNA.  
 XX  
 KW Mouse; transcription factor; neurogenin 3; ngn3; stem cell;  
 KW differentiation; beta-cell; insulin; diabetes; hyperglycaemia;  
 KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;  
 KW gene; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 160..804  
 FT /\*tag= a

FT /product= "Mouse ngn3"  
 XX  
 XX WO200286107-A2.  
 XX  
 XX 31-OCT-2002.  
 PD  
 XX  
 PF 19-APR-2002; 2002WO-EP04362.  
 XX  
 PR 19-APR-2001; 2001US-284531P.  
 XX  
 PA (DEVE-) DEVELOPEN ENTWICKLUNGSBIOLOGISCHE FORSCH.  
 PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.  
 XX  
 PI Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;  
 XX  
 DR MPI; 2003-075629/07.  
 XX  
 XX  
 PT Differentiating stem cells into insulin-producing cells useful for  
 PT treating pancreatic diseases, by culturing stem cells in suitable  
 PT medium and activating gene involved in beta-cell differentiation -  
 XX  
 PS Disclosure; Page 58-59; 62pp; English.  
 XX  
 CC The present sequence is that of cDNA encoding the murine  
 CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),  
 CC a gene which is required for the specification of the early  
 CC endocrine precursor in the pancreatic epithelium and which is  
 CC down-regulated once endocrine differentiation begins. The invention  
 CC provides a claimed method for differentiating stem cells (especially  
 CC embryonic, adult or somatic stem cells and primordial germ cells)  
 CC into insulin-producing cells. This involves culturing stem cells in  
 CC a suitable medium and activating at least one gene involved in  
 CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,  
 CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the  
 CC delivery of the gene into stem cells using a viral delivery  
 CC system, or the delivery of a protein product of the gene into stem  
 CC cells. The insulin-producing cells can be transplanted into  
 CC animals or human for treatment of pancreatic diseases, metabolic  
 CC syndrome and metabolic disorders with impaired glucose levels such  
 CC as diabetes, hyperglycaemia and impaired glucose tolerance  
 CC (claimed). The cells can also be used to identify compounds which  
 CC stimulate beta-cell differentiation, insulin secretion or glucose  
 CC responsiveness. Differentiated beta-cells can be used to study the  
 CC toxic and other effects of exogenous compounds on beta-cell  
 CC function. In an example from the invention, Pax6 cDNA was inserted  
 CC into expression vector pACCMV.plpa under the control of the  
 CC cytomegalovirus promoter.  
 XX  
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;  
 QY  
 Query Match 6.3%; Score 92; DB 25; Length 861;  
 Best Local Similarity 100.0%; Pred. No. 9e-35;  
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 762 GGGCTGATGCGCTGCGGCTGCTCCGACCTTCCGGATGAGCGCAACTTACAAG 821  
 463 GGGCTGATGCGCTGCGGCTGCTCCGACCTTCCGGATGAGCGCAACTTACAAG 522  
 QY 822 ATCGAGACCTTGCGCTTCCGCCCAACTACAT 853  
 Db 523 ATCGAGACCTTGCGCTTCCGCCCAACTACAT 554  
 RESULT 6  
 AAD46872  
 ID AAD46872 strand; DNA; 1860 BP.  
 AC AAD46872;  
 XX  
 DT 27-JAN-2003 (first entry)  
 DE Murine neurogenin 3 (Ngn3) gene.  
 XX

KM	Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;	
KW	type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;	
KX	islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.	
XX		
OS	Mus musculus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1093..1737
FT		/*tag= a
FT		/product= "Murine Ngn3 protein"
XX		
PN	WO200274045-A2.	
XX		
PD	26-SEP-2002.	
XX		
PP	20-MAR-2002; 2002WO-US11166.	
XX		
PR	20-MAR-2001; 2001US-0817360.	
XX		
PA	(REGC ) UNIV CALIFORNIA.	
XX		
PI	German MS, Lin J;	
XX		
DR	WPI; 2002-759853/82.	
DR	P-PSDB; AAE29278.	
XX		
PT	Producing a mammalian islet cell for treating diabetes mellitus	
PT	comprises introducing into a mammalian cell a nucleic acid molecule	
XX	encoding neuroendocrine basic helix-loop-helix transcription factor	-
PS	Example 3; Page 89-90; 108pp; English.	
XX		
CC	The invention relates to a method for producing a mammalian islet cell.	
CC	The method comprising introducing into a mammalian cell a nucleic acid	
CC	molecule encoding an islet transcription factor for expression of the	
CC	islet transcription factor in the cell and for production of islet cell	
CC	phenotype in the cell. The islet transcription factor is a neuroendocrine	
CC	basic helix-loop-helix (bHLH) transcription factor. The method is useful	
CC	for treating type 2 diabetes mellitus and for replacing beta cells lost	
CC	to autoimmune destruction in individuals with type 1 diabetes. The method	
CC	is useful in cell therapy. The present sequence is murine neurogenin 3	
CC	(Ngn3) gene.	
XX		
SQ	Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;	
XX		
Query Match	6.3%; Score 92; DB 24; Length 1860;	
Best Local Similarity	100.0%; Pred. No. 8.7e-35;	
Matches	92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	762 GCGCTGATGCGCTGCGGCGGTGCTGCCACCTTCGCGATGACGCCAACTTACAAAG 821	
DB	1396 GCGCTGATGCGCTGCGGCGGTGCTGCCACCTTCGCGATGACGCCAACTTACAAAG 1455	
QY	822 ATCGAGACCTGCGCTTGCGCCACACTACAT 853	
DB	1456 ATCGAGACCTGCGCTTGCGCCACACTACAT 1487	
XX		
RESULT 7		
ACAC61090		
ID	ACAC61090 standard; DNA; 1861 BP.	
XX		
AC	ACAC61090;	
XX		
DT	05-FEB-2001 (first entry)	
XX		
DE	Murine neurogenin 3 (Ngn3) genomic DNA sequence.	
XX		
XX	Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;	
KW	islet cell precursor identification; mouse; ds.	
XX		
OS	Mus musculus.	
XX		

[illegible]

```

XX OS Mus musculus.
XX PN WO200073764-A2.
XX PD 07-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US15410.
XX PR 01-JUN-1999; 99US-0137060.
XX PR 19-JAN-2000; 2000US-0176993.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PI Zoghbi HY, Bellien H, Birmingham N, Haasan B, Ben-Arie N;
XX DR WPI, 2001-032190/04.
XX DR P-PSDB; AAB60350.
XX PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
XX PT any of its homologs or orthologs, for the treatment of e.g. deafness,
XX PT osteoarthritis and abnormal cell proliferation -
XX PS Disclosure; Page -: 142pp; English.
XX CC The invention relates to the use of atonal-associated nucleic acid or
XX CC amino acid sequence, or any of its homologues or orthologues as
XX CC therapeutic agents for the treatment of deafness, partial hearing loss,
XX CC vestibular effects due to damage or loss of inner hair cells,
XX CC osteoarthritis and abnormal cell proliferation. The invention also
XX CC encompasses methods of screening for compounds which affect the
XX CC expression of an atonal-associated nucleic acid sequence in an animal,
XX CC and a transgenic animal in which an allele of a native atonal-associated
XX CC gene is replaced by a heterologous nucleic acid sequence, thus
XX CC inactivating the atonal-associated allele. The nucleic acids or proteins
XX CC may be used in a method of treating an animal for hearing impairment,
XX CC joint disease, balance disorders, abnormal cell proliferation, or other
XX CC disease related to loss of a functional atonal-associated nucleic acid or
XX CC protein. They may particularly be used to treat an animal with a
XX CC deficiency in cerebellar granule neurons or their precursors, and may
XX CC also be used in promoting mechanoreceptive cell growth and generating
XX CC hair cells. The present sequence represents an atonal-associated nucleic
XX CC acid sequence referred to in the invention.
XX CC Note: The present sequence is not shown in the specification, but
XX CC was obtained from Genbank.
XX SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
XX
XX Query Match 6.3%; Score 92; DB 22; Length 5567;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-35;
XX Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 762 GCGGTGATGCGCTGCGGCTGCTCTGCGCCACCTTCCGGATGAGCGCAACTTACAAAG 821
DB 5226 GCGGTGATGCGCTGCGGCTGCTCTGCGCCACCTTCCGGATGAGCGCAACTTACAAAG 5285
OY 822 ATCGAGACCCCTGCGCTTCCGCCCAACTCAT 853
DB 5286 ATCGAGACCCCTGCGCTTCCGCCCAACTCAT 5317

```

```

XX OS Rattus norvegicus.
XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-1B01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287722A.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Fatgler S;
XX DR WPI, 2002-257383/30.
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX PS Example 1; SEQ ID 4140; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridizing selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterizing the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIP0
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;
XX
XX Query Match 2.5%; Score 36; DB 24; Length 65;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-07;
XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1281 ATTGAGGCTATCTCCTTAACCCCTCCTAGTGT 1316
DB 30 ATTGAGGCTATCTCCTTAACCCCTCCTAGTGT 65

```

```

RESULT 9
ABN31392
ID ABN31392 standard; DNA; 65 BP.
XX AC ABN31392;
XX XX
XX XX
XX DT 15-JUL-2002 (first entry)
XX DE Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.
XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

```

```

RESULT 10
ABN57521
ID ABN57521 standard; DNA; 65 BP.
XX AC ABN57521;
XX XX
XX XX
XX DT 15-JUL-2002 (first entry)
XX DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.
XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

```





/note= "Neurogenin 3"

FT XX WO20005936-A1.  
XX XX 12-OCT-2000.  
XX PD 28-MAR-2000; 2000WO-US08436.  
XX PF 06-APR-1999; 99US-0128180.  
XX PR (REGC ) UNIV CALIFORNIA.  
XX PA German MS, Lin J;  
XX PI WPI; 2000-664989/64.  
XX DR P-PSDB; AAY85617.  
XX DR Novel human neurogenin 3 polypeptides and polynucleotides encoding  
PT them, useful for diagnosis, prevention and treatment of diabetes  
PT mellitus and to identify individuals at risk of diabetes -  
XX PS Claim 6; Page 46-48; 54pp; English.  
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3  
CC protein AAY85617. The Ngn3 gene is located at chromosome position  
CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and  
CC protein sequences, and includes an antibody recognising the Ngn3 protein.  
CC Also included in the invention is a method for identifying an islet cell  
CC precursor, the method involves analysing a cell for the expression of the  
CC Ngn3 gene product, where detection of the product is indicative of an  
CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic  
CC reagent for detecting (in a subject) a predisposition to a defect in  
CC pancreatic islet cell function or formation associated with a defect in  
CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell  
CC precursor cells expressing Ngn3, and to alter cellular differentiation in  
CC culture in vivo to produce new beta-cells to treat patients with diabetes  
CC mellitus.  
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;  
Query Match 2.2%; Score 32; DB 21; Length 5340;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 672 AAGAGCGAGTTGGCACTGACGACGAGCGAGC 703  
DB 3235 AAGAGCGAGTTGGCACTGACGACGAGCGAGC 3266  
RESULT 13  
AAD46871  
ID AAD46871 standard; DNA; 5340 BP.  
XX AC AAD46871;  
XX DT 27-JAN-2003 (first entry)  
XX DE Human neurogenin 3 (Ngn3) gene.  
XX KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;  
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;  
KW islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;  
KW gene; ds.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT CDS 3022..3666  
FT /tag= a  
FT /product= "Human Ngn3 protein"  
XX WO200274045-A2.

PD 26-SEP-2002.  
XX PF 20-MAR-2002; 2002WO-US11166.  
XX PR 20-MAR-2001; 2001US-0817360.  
XX PA (REGC ) UNIV CALIFORNIA.  
XX PI German MS, Lin J;  
XX DR WPI; 2002-759853/82.  
XX DR P-PSDB; AAE29277.  
XX PT Producing a mammalian islet cell for treating diabetes mellitus  
PT comprises introducing into a mammalian cell a nucleic acid molecule  
PT encoding neuroendocrine basic helix-loop-helix transcription factor -  
XX PS Example 2; Page 87-88; 108pp; English.  
XX CC The invention relates to a method for producing a mammalian islet cell.  
XX CC The method comprising introducing into a mammalian cell a nucleic acid  
XX CC molecule encoding an islet transcription factor for expression of the  
XX CC islet transcription factor in the cell and for production of islet cell  
XX CC phenotype in the cell. The islet transcription factor is a neuroendocrine  
XX CC basic helix-loop-helix (bHLH) transcription factor. The method is useful  
XX CC for treating type 2 diabetes mellitus and for replacing beta cells lost  
XX CC to autoimmune destruction in individuals with type 1 diabetes. The method  
XX CC is useful in cell therapy. The present sequence is human neurogenin 3  
XX CC (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.  
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;  
Query Match 2.2%; Score 32; DB 24; Length 5340;  
Best Local Similarity 100.0%; Pred. No. 2.4e-05;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 672 AAGAGCGAGTTGGCACTGACGACGAGCGAGC 703  
DB 3235 AAGAGCGAGTTGGCACTGACGACGAGCGAGC 3266  
RESULT 14  
AAS33797  
ID AAS33797 standard; CDNA; 428 BP.  
XX AC AAS33797;  
XX DT 17-DEC-2001 (first entry)  
XX DE Human cDNA encoding a novel foetal antigen, SEQ ID No 321.  
XX KW Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;  
KW immunomodulator; cardiovascular; cytosolic; nephroprotective;  
KW cardiovascular; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; breast neoplasm; cancer;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; food additive.  
XX OS Homo sapiens.  
XX PN WO200155312-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01321.  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-020515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235844.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-488782/53.  
DR P-PSDB; AAU20977.  
XX  
XX  
PT New polynucleotides and polypeptides for diagnosing, treating,  
PT preventing or prognosing e.g. diseases or disorders of the nervous,  
PT musculoskeletal, excretory, gastrointestinal, reproductive, and  
PT respiratory systems -  
XX  
XX  
PS Claim 1; SEQ ID No 321, 642pp; English.  
XX

CC The invention relates to novel nucleic acids encoding novel human foetal  
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.  
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. The antibodies to the antigens can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. Numerous  
CC examples of diseases and disorders treated by the nucleic acids and  
CC proteins are given in the specification. The present sequence

Query Match 2.0%; Score 29; DB 22; Length 428;  
Best Local Similarity 100.0%; Pred. No. 0.0008;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 825 GAGACCTGCGCTGCGCCCAACAACATACAT 853  
DB 1 GAGACCTGCGCTGCGCCCAACAACATACAT 29

RESULT 15  
ABT03700/C  
ID ABT03700 standard; DNA; 26 BP.

XX AC ABT03700;

XX DT 13-SEP-2002 (first entry)

XX DE Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.

XX KW Human; cancer; neoplastic disease; tumour specific marker; cytostatic;  
XX transcription factor; PCR; primer; ss.

XX OS Homo sapiens.

XX PN WO200240716-A2.

XX PD 23-MAY-2002.

XX PF 13-NOV-2001; 2001WO-US43461.

XX PR 16-NOV-2000; 2000US-249508P.

XX PA (CEMI-) CEMINES LLC.

XX PI Palm K;

XX DR WPI; 2002-537346/57.

PT Determining the presence of neoplastic molecular markers, by  
PT identifying the presence of markers in host test sample using array of  
PT neoplastic molecular marker specific reagents and analyzing the array  
PT of the reagents -

XX PS Example 1; Page 17; 41pp; English.

CC The present invention relates to a method for determining the presence of  
CC neoplastic molecular markers in a host, involving the use of neoplastic  
CC molecular marker specific reagents to detect such markers and analyzing

CC the array of reagents, allowing the identification of the neoplastic  
CC disease present. This can be used to determine the best treatment for  
CC cancer, in particular neural cell, lung and prostate tumours. The  
CC present sequence is a PCR primer useful for detecting the coding  
CC sequences of markers of the invention.

XX SQ Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;

Query Match 1.8%; Score 26; DB 24; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.027;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 673 AGAGCACTGCGCACTGAGCAAGCAG 698  
DB 26 AGAGCACTGCGCACTGAGCAAGCAG 1

RESULT 16  
AAV42515  
ID AAV42515 standard; DNA; 25 BP.

XX AC AAV42515;

XX DT 05-OCT-1998 (first entry)

XX DE PCR primer used to isolate part of the RELAX protein coding region.

XX KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;  
XX control; gene expression; transcriptional activator; targeting;  
XX protein expression; central nervous system; CNS; treatment;  
XX nervous system disorder; CIG235; PCR primer; ss.

XX OS Synthetic.

XX OS Rattus sp.

XX PN WO9827206-A2.

XX PD 25-JUN-1998.

XX PF 19-DEC-1997; 97WO-FR02368.

XX PR 19-DEC-1996; 96FR-0015651.

XX PA (RHON ) RHONE-POULENC RORER SA.

XX PI Mallet J, Ravassard P, Icard-Liepkalns C;

XX DR WPI; 1998-362775/31.

PT Basic helix-loop-helix polypeptide and related nucleic acid - with  
PT transcriptional activity, for targeting expression of genes to  
PT central nervous system and treatment of nervous disease

XX PS Example 2; Page 12; 28pp; French.

CC PCR primers AAV42515-16 are used to isolate part of the DNA encoding  
CC a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat  
CC Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.  
CC The protein is used to control and participate in gene expression,  
CC by acting as transcriptional activator, strictly dependent on the  
CC presence of an intact E box (CANNTG), particularly for targeting  
CC expression of proteins to the central nervous system (CNS). The  
CC nucleic acid sequence can be used to treat nervous system disorders,  
CC and antisense sequences can be used to control mRNA transcription.

XX SQ Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.083;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 750 AACCTTAACCTGCGCGCTGATGCGC 774  
|||||

Db 1 AACCTTAACCGCGCTGATCGGC 25

RESULT 17  
AAV27046  
ID AAV27046 standard; CDNA, 738 BP.  
XX  
AC AAV27046;  
XX  
DT 17-SEP-1998 (first entry)  
XX  
DE Mouse neurogenin 1 gene.  
XX  
KW ss; Mouse; neurogenin; expression vector; recombinant protein;  
antibody; neurogenesis.  
XX  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..735  
FT /tag= a  
FT /product= "Mouse neurogenin 1"  
XX  
XX WO9813491-A2.  
XX  
XX 02-APR-1998.  
XX  
XX 24-SEP-1997; 97WO-US17048.  
XX  
XX 17-SEP-1997; 97US-0932411.  
XX 27-SEP-1996; 96US-0722570.  
XX 12-NOV-1996; 96US-0030864.  
XX 19-DEC-1996; 96US-0772009.  
XX  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
XX  
XX Anderson DJ, Ma Q, Sommer L;  
XX  
XX MPI, 1998-230702/20.  
XX  
XX P-PSDB; AAW54944.  
XX  
XX Mouse neurogenin, useful in neurogenesis - and recombinant nucleic  
PT acids and proteins derived from rat and Xenopus  
XX  
XX Claim 5; Fig 4; 106pp; English.  
XX  
XX The mouse neurogenin 1 is one of several neurogenin proteins discussed in  
CC the present invention. The neurogenin nucleic acids can be expressed in  
CC a host cell, transformed using an expression vector, to produce  
CC recombinant proteins. The proteins and the antibodies raised against  
CC the proteins are useful in the study of neurogenesis.  
XX  
XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;  
SQ

Query Match 1.6%; Score 23; DB 19; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.69;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754  
DB 304 GAGCGCAACCGCATGCACAACT 326

RESULT 18  
AAZ51977  
ID AAZ51977 standard; DNA; 738 BP.  
XX  
AC AAZ51977;  
XX  
DT 04-JUL-2000 (first entry)  
XX  
DE Murine neurogenin-1 (NGN1) nucleic acid sequence.  
XX

KW Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;  
KW Phox2a protein; neuronal subtype-specific marker; growth factor;  
KW neural differentiation; transplantantion; neuronal dysfunction;  
KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;  
KW neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;  
KW cerebroprotective; immunosuppressant; antiinfectious; ss.  
XX  
XX Mus sp.  
XX  
XX Key Location/Qualifiers  
FT CDS 1..735  
FT /tag= a  
FT /product= "Murine neurogenin-1 protein"  
XX  
XX WO200009676-A2.  
XX  
XX 24-FEB-2000.  
XX  
XX 13-AUG-1999; 99WO-US18525.  
XX  
XX 14-AUG-1998; 98US-0096630.  
XX  
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.  
XX  
XX Anderson DJ, Lo L;  
XX  
XX MPI, 2000-256250/22.  
XX P-PSDB; AAY70566.  
XX  
XX Inducing non-neuronal cells to differentiate into neurons and for  
PT non-neuronal cells to express a neuronal subtype-specific marker,  
PT comprising contacting the non-neuronal cells with a vector containing  
PT neurogenin nucleic acid -  
XX  
XX Claim 1; Fig 1C; 76pp; English.  
XX  
XX The patent discloses a method for inducing non-neuronal cells (NNC) to  
CC differentiate into neurons and for NNCs to express a neuronal subtype  
CC -specific marker. Transformed host cells are used as sources of neuronal  
CC and other growth factors; in culture for screening compounds that  
CC modulate neural differentiation or as sources of recombinantly produced  
CC neurogenins and Phox2a proteins for use in transplantation. The cells  
CC also have a variety of in vivo uses, e.g. for transplantation at sites of  
CC neuronal dysfunction e.g. patients with hearing or vision loss due to  
CC optical or auditory nerve damage, brain or spinal cord injuries, and  
CC neurodegenerative disorders e.g. Alzheimer's disease. The present  
CC sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.  
CC NNCs differentiate into neurons through the recombinant expression of a  
CC transcription factor that induces a core program of neurogenesis. Forced  
CC expression of murine NGN1 can elicit expression of at least some neuronal  
CC phenotypic markers even in NNCs. This can be used in autografting.  
XX  
XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;  
SQ

Query Match 1.6%; Score 23; DB 21; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.69;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754  
DB 304 GAGCGCAACCGCATGCACAACT 326

RESULT 19  
AAF27264  
ID AAF27264 standard; CDNA; 790 BP.  
XX  
AC AAF27264;  
XX  
DT 24-APR-2001 (first entry)  
XX  
DE Chicken atonal homologue ngn2/ath4a CDNA, SEQ ID NO:20.  
XX

KW Atonal; homologue; orthologue; atonal-associated protein; deafness;  
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
 KW cellular proliferation; cerebellar granule neuron; gene therapy;  
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
 KW transgenic animal; ss.  
 OS Gallus gallus.  
 XX  
 XX WO200073764-A2.  
 PN  
 XX  
 XX 07-DEC-2000.  
 PD  
 XX  
 XX 01-JUN-2000; 2000WO-US15410.  
 PF  
 XX  
 XX 01-JUN-1999; 99US-0137060.  
 PR  
 XX 19-JAN-2000; 2000US-0176993.  
 PR  
 XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
 PA  
 XX  
 XX Zoghbi HY, Bellen H, Birmingham N, Haasan B, Ben-Arie N;  
 PI WPI; 2001-032190/04.  
 XX P-PSDB; AAB60357.  
 DR  
 XX  
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,  
 PT osteoarthritis and abnormal cell proliferation -  
 PT  
 XX Disclosure; Page -, 142pp; English.  
 PS  
 XX The invention relates to the use of atonal-associated nucleic acid or  
 CC amino acid sequence, or any of its homologs or orthologues as  
 CC therapeutic agents for the treatment of deafness, partial hearing loss,  
 CC vestibular effects due to damage or loss of inner hair cells,  
 CC osteoarthritis and abnormal cell proliferation. The invention also  
 CC encompasses methods of screening for compounds which affect the  
 CC expression of an atonal-associated nucleic acid sequence in an animal,  
 CC and a transgenic animal in which an allele of a native atonal-associated  
 CC gene is replaced by a heterologous nucleic acid sequence, thus  
 CC inactivating the atonal-associated allele. The nucleic acids or proteins  
 CC may be used in a method of treating an animal for hearing impairment,  
 CC joint disease, balance disorders, abnormal cell proliferation, or other  
 CC disease related to loss of a functional atonal-associated nucleic acid or  
 CC protein. They may particularly be used to treat an animal with a  
 CC deficiency in cerebellar granule neurons or their precursors, and may  
 CC also be used in promoting mechanoreceptive cell growth and generating  
 CC hair cells. The present sequence represents an atonal-associated nucleic  
 CC acid sequence referred to in the invention.  
 CC Note: The present sequence is not shown in the specification, but  
 CC was obtained from GenBank.  
 CC  
 XX  
 SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;  
 Query Match 1.6%; Score 23; DB 22; Length 790;  
 Best Local Similarity 100.0%; Pred. No. 0.69;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 732 GAGCGCAACCGCATGCACACCT 754  
 Db 374 GAGCGCAACCGCATGCACACCT 396  
 RESULT 20  
 ID AAV42938 standard; DNA; 1332 BP.  
 AC AAV42938;  
 XX  
 XX 25-MAR-2003 (updated)  
 DT 21-OCT-1998 (first entry)  
 XX  
 XX DNA encoding murine neuroD3 protein, which is a bHLH protein.  
 DE  
 XX

KW Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;  
 KW classification; medulloblastoma; mouse; ds.  
 XX  
 XX Mus musculus.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH CDS 101..835  
 FT /\*tag= a  
 FT /product= neuroD3  
 FT  
 XX  
 XX US5795723-A.  
 PN  
 XX  
 XX 18-AUG-1998.  
 PD  
 XX  
 XX 07-AUG-1997; 97US-0910973.  
 PF  
 XX  
 XX 06-MAY-1994; 94US-0239238.  
 PR  
 XX 02-NOV-1995; 95US-0552142.  
 PR 30-OCT-1996; 96WO-US17532.  
 XX  
 XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 PA  
 XX  
 XX Olson JM, Tapscott SJ;  
 PI WPI; 1998-466661/40.  
 XX P-PSDB; AAW71019.  
 DR  
 XX  
 XX Classifying neuroectodermal tumours from expression pattern of  
 PT basic-helix-loop-helix genes - especially for identifying  
 PT medulla:blastoma and assessing its aggressiveness, specifically  
 PT associated with expression of bHLH genes neuroD 1-3  
 PT  
 XX Example 11; Columns 75-78; 43pp; English.  
 PS  
 XX The present sequence encodes a protein which is a member of the basic  
 CC helix-loop-helix (bHLH) protein family, and is designated neuroD3. The  
 CC bHLH genes are a family of genes associated with vertebrate neuronal,  
 CC endocrinal and gastrointestinal development. The observed pattern of  
 CC neuroD expression distinguishes subclasses of neuroectodermal tumours.  
 CC The specification describes a method for the classification of human  
 CC neuroectodermal tumours. The method comprises measuring, in a tumour  
 CC sample, expression of at least one basic bHLH gene and identifying the  
 CC tumour subclass by matching expression to predetermined expression  
 CC profiles for known subclasses. For classifying the tumour as a  
 CC medulloblastoma, the bHLH gene detected is neuroD1 and neuroD3.  
 CC The method is used to classify neuroectodermal tumours, and to identify  
 CC medulloblastoma and for prognosis of this as aggressive.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 CC  
 XX  
 SQ Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;  
 Query Match 1.6%; Score 23; DB 19; Length 1332;  
 Best Local Similarity 100.0%; Pred. No. 0.67;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 732 GAGCGCAACCGCATGCACACCT 754  
 Db 404 GAGCGCAACCGCATGCACACCT 426  
 RESULT 21  
 ID ABS56396 standard; DNA; 1332 BP.  
 AC ABS56396;  
 XX  
 XX 23-JAN-2003 (first entry)  
 DT  
 XX  
 XX Mouse bHLH family neuroD3 genomic DNA.  
 DE  
 XX Mouse; gene; ds; neuroD3; neuroD; basic-helix-loop-helix; bHLH;  
 KW differentiation; neurone; endocrine; gastrointestinal; development;  
 KW transgenic; embryo; birth defect; spontaneous abortion; stem cell;  
 KW

KW cancer;neural growth factor; tumour; diagnostic; motor; sensory;  
 KM traumatic neural injury; hearing; vision; brain; spinal cord;  
 KM malabsorption syndrome; gastrointestinal dysmotility syndrome;  
 KM Hirsh Prung's disease; therapeutic.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 101..835  
 FT /tag= a  
 FT /product= "NeuroD3"  
 FT misc\_feature 425..544  
 FT /tag= b  
 FT /note= "HLH coding domain"  
 XX  
 XX US6444463-B1.  
 XX  
 XX 03-SEP-2002.  
 XX  
 PD 07-FEB-2000; 2000US-0499227.  
 PF  
 PR 05-AUG-1998; 98WO-US16417.  
 PR 06-MAY-1994; 94US-0239238.  
 PR 08-MAY-1995; 95WO-US05741.  
 PR 02-NOV-1995; 95US-0552142.  
 PR 30-OCT-1996; 96WO-US17532.  
 PR 07-AUG-1997; 97US-0910973.  
 PA  
 XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
 XX  
 PI Tapscott SJ;  
 XX  
 XX WPI; 2003-056678/05.  
 DR P-PSDB; ABG72005.  
 XX  
 PT New neurogenic differentiation gene, useful in gene therapy to correct  
 PT traumatic neural injury that has resulted in loss of motor or sensory  
 PT neural function and for constructing recombinant cell lines -  
 XX  
 XX Example 11; Column 75-78; 43pp; English.  
 PS  
 XX  
 CC The invention discloses an isolated nucleic acid molecule which encodes a  
 CC functionally active human neuroD3 polypeptide. NeuroD proteins represent  
 CC a new family within the basic-helix-loop-helix (bHLH) family which are  
 CC implicated in the regulation of differentiation. NeuroD proteins are  
 CC particularly involved in neuronal, endocrine and gastrointestinal  
 CC development. The nucleic acid is useful for constructing recombinant cell  
 CC lines, transgenic embryos and animals and for quantifying the level of  
 CC expression of neuroD in a cell. Birth defects and spontaneous abortions  
 CC may result from expression of an abnormal neuroD protein. The  
 CC polynucleotide sequences permit the establishment of primary cultures of  
 CC proliferating embryonic neuronal stem cells under conditions mimicking  
 CC those that are active in development and cancer. The resultant cell lines  
 CC find use as sources of novel neural growth factors, in assays for  
 CC identifying novel neuronal growth factors which can be used for screening  
 CC anti-cancer drugs capable of driving terminal differentiation in neural  
 CC tumours, for producing antibodies useful in diagnostic assays and for  
 CC screening for compounds capable of modulating the activity of neuroD.  
 CC Transformed host cells, nucleic acids and polypeptides are also useful  
 CC for treating sites of traumatic neural injury where motor or sensory  
 CC neural activity has been lost, e.g. hearing or vision loss and brain or  
 CC spinal cord damage. The host cells find use in the treatment of  
 CC malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsh  
 CC Prung's disease). The cell lines also find use in screening for candidate  
 CC therapeutic agents capable of either substituting for neuroD or  
 CC correcting the cellular defect caused by a defective neuroD. The sequence  
 CC presented is the mouse neuroD3 genomic DNA.  
 XX  
 XX Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;  
 SQ  
 Query Match 1.6%; Score 23; DB 25; Length 1332;  
 Best Local Similarity 100.0%; Pred. No. 0.67; Mismatches 23; Conservative 0; Indels 0; Gaps 0

[illegible]

```

RESULT 23
AAV27049
ID AAV27049 standard; cDNA; 1385 BP.
XX
XX AAV27049;
XX
XX 17-SEP-1998 (first entry)
XX
XX Mouse neurogenin 2 gene.
DE
XX
XX ds; Mouse; neurogenin; expression vector; recombinant protein;
XX antibody; neurogenesis.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX CDS 382..1173
XX /*tag= a
XX /product= "Mouse neurogenin 2"

MO9813491-A2.
XX
XX 02-APR-1998.
XX
XX 24-SEP-1997; 97WO-US17048.
XX
XX 17-SEP-1997; 97US-0932411.
XX 27-SEP-1996; 96US-0722570.
XX 12-NOV-1996; 96US-0030864.
XX 19-DEC-1996; 96US-0772009.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Anderson DJ, Ma Q, Sommer L;
XX
XX WPI; 1998-230702/20.
XX
XX P-PSDB; AAW54946.
XX
XX Mouse neurogenin, useful in neurogenesis - and recombinant nucleic
XX acids and proteins derived from rat and Xenopus
XX
XX Claim 5; Fig 7; 106pp; English.
XX
XX The Mouse neurogenin 2 is one of several neurogenin proteins discussed
XX in the present invention. The neurogenin nucleic acids can be expressed
XX in a host cell, transformed using an expression vector, to produce
XX recombinant proteins. The proteins and the antibodies raised against
XX the proteins are useful in the study of neurogenesis.
XX
XX Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;
SQ

Query Match 1.6%; Score 23; DB 19; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 732 GAGCGCAACCGCATGCACAACCT 754
Db 742 GAGCGCAACCGCATGCACAACCT 764

```

```

KW optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KW neuroprotective; neurotrophic; anticonvulsant; antiParkinsonian; vulnerary;
KW cerebroprotective; immunosuppressant; antiinfectious; ds.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX CDS 382..1173
XX /*tag= a
XX /product= "Murine neurogenin-2 protein"

WO200009676-A2.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18525.
XX
XX 14-AUG-1998; 98US-0096630.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Anderson DJ, Lo L;
XX
XX WPI; 2000-256250/22.
XX
XX P-PSDB; AAV70569.
XX
XX Inducing non-neuronal cells to differentiate into neurons and for
XX non-neuronal cells to express a neuronal subtype-specific marker,
XX comprising contacting the non-neuronal cells with a vector containing
XX neurogenin nucleic acid -
XX
XX Claim 1; Fig 11; 76pp; English.
XX
XX The patent discloses a method for inducing non-neuronal cells (NNC) to
XX differentiate into neurons and for NNCs to express a neuronal subtype
XX -specific marker. Transformed host cells are used as sources of neuronal
XX and other growth factors; in culture for screening compounds that
XX modulate neural differentiation or as sources of recombinantly produced
XX neurogenins and Phox2a proteins for use in transplantation. The cells
XX also have a variety of in vivo uses, e.g. for transplantation at sites of
XX neuronal dysfunction e.g. patients with hearing or vision loss due to
XX optical or auditory nerve damage, brain or spinal cord injuries, and
XX neurodegenerative disorders e.g. Alzheimer's disease. The present
XX sequence encodes murine neurogenin-2 (NGN-2), a transcription factor.
XX NNCs differentiate into neurons through the recombinant expression of a
XX transcription factor that induces a core program of neurogenesis. Forced
XX expression of murine NGN2 can elicit expression of at least some neuronal
XX phenotypic markers even in NNCs.
XX
XX Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;
SQ

Query Match 1.6%; Score 23; DB 21; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 732 GAGCGCAACCGCATGCACAACCT 754
Db 742 GAGCGCAACCGCATGCACAACCT 764

```

```

RESULT 24
AAZ51980
ID AAZ51980 standard; DNA; 1385 BP.
XX
XX AAZ51980;
XX
XX 04-JUL-2000 (first entry)
XX
XX Murine neurogenin-2 (NGN-2) nucleic acid sequence.
XX
XX Neurogenin-2; NGN-2; non-neuronal cell; NNC; neurogenesis;
XX Phox2a protein; neuronal subtype-specific marker; growth factor;
XX neural differentiation; transplantation; neuronal dysfunction;
KW

```

```

RESULT 25
AAF27269
ID AAF27269 standard; cDNA; 1385 BP.
XX
XX AAF27269;
XX
XX 24-APR-2001 (first entry)
XX
XX Mouse neurogenin 2 (ngn2) cDNA, SEQ ID NO:30.
XX
XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW

```



KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
KW transgenic animal; ss.  
XX Mus musculus.  
XX  
XX WO200073764-A2.  
XX  
XX 07-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15410.  
XX  
XX 01-JUN-1999; 99US-0137060.  
XX  
XX 19-JAN-2000; 2000US-0176993.  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
XX  
XX WPI; 2001-032190/04.  
XX  
XX P-PSDB; AAB60362.  
XX  
XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
XX any of its homologs or orthologs, for the treatment of e.g. deafness,  
XX osteoarthritis and abnormal cell proliferation -  
XX  
XX Disclosure; Page -: 142pp; English.  
XX  
XX The invention relates to the use of atonal-associated nucleic acid or  
XX amino acid sequence, or any of its homologues or orthologues as  
XX therapeutic agents for the treatment of deafness, partial hearing loss,  
XX vestibular effects due to damage or loss of inner hair cells,  
XX osteoarthritis and abnormal cell proliferation. The invention also  
XX encompasses methods of screening for compounds which affect the  
XX expression of an atonal-associated nucleic acid sequence in an animal,  
XX and a transgenic animal in which an allele of a native atonal-associated  
XX gene is replaced by a heterologous nucleic acid sequence, thus  
XX inactivating the atonal-associated allele. The nucleic acids or proteins  
XX may be used in a method of treating an animal for hearing impairment,  
XX joint disease, balance disorders, abnormal cell proliferation, or other  
XX disease related to loss of a functional atonal-associated nucleic acid or  
XX protein. They may particularly be used to treat an animal with a  
XX deficiency in cerebellar granule neurons or their precursors, and may  
XX also be used in promoting mechanoreceptive cell growth and generating  
XX hair cells. The present sequence represents an atonal-associated nucleic  
XX acid sequence referred to in the invention.  
XX Note: The present sequence is not shown in the specification, but  
XX was obtained from GenBank.  
XX  
XX Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;  
XX  
XX Query Match 1.6%; Score 23; DB 22; Length 1385;  
XX Best Local Similarity 100.0%; Pred. No. 0.67;  
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGGCAACCGCATGCACAACCT 754  
Db 742 GAGGCAACCGCATGCACAACCT 764

RESULT 26  
AAF27255  
ID AAF27255 standard; cDNA; 1412 BP.  
XX  
XX AAF27255;  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX Mouse atonal homologue 4 (ATOH4, Math4A) cDNA, SEQ ID NO:6.  
XX  
XX Atonal; homologue; orthologue; atonal-associated protein; deafness;  
XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
XX cellular proliferation; cerebellar granule neuron; gene therapy;  
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;

KW transgenic animal; ss.  
XX  
XX Mus musculus.  
XX  
XX WO200073764-A2.  
XX  
XX 07-DEC-2000.  
XX  
XX 01-JUN-2000; 2000WO-US15410.  
XX  
XX 01-JUN-1999; 99US-0137060.  
XX  
XX 19-JAN-2000; 2000US-0176993.  
XX  
XX (BAYU ) BAYLOR COLLEGE MEDICINE.  
XX  
XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;  
XX  
XX WPI; 2001-032190/04.  
XX  
XX P-PSDB; AAB60351.  
XX  
XX Therapeutic use of atonal-associated nucleic acids or amino acids, or  
XX any of its homologs or orthologs, for the treatment of e.g. deafness,  
XX osteoarthritis and abnormal cell proliferation -  
XX  
XX Disclosure; Page -: 142pp; English.  
XX  
XX The invention relates to the use of atonal-associated nucleic acid or  
XX amino acid sequence, or any of its homologues or orthologues as  
XX therapeutic agents for the treatment of deafness, partial hearing loss,  
XX vestibular effects due to damage or loss of inner hair cells,  
XX osteoarthritis and abnormal cell proliferation. The invention also  
XX encompasses methods of screening for compounds which affect the  
XX expression of an atonal-associated nucleic acid sequence in an animal,  
XX and a transgenic animal in which an allele of a native atonal-associated  
XX gene is replaced by a heterologous nucleic acid sequence, thus  
XX inactivating the atonal-associated allele. The nucleic acids or proteins  
XX may be used in a method of treating an animal for hearing impairment,  
XX joint disease, balance disorders, abnormal cell proliferation, or other  
XX disease related to loss of a functional atonal-associated nucleic acid or  
XX protein. They may particularly be used to treat an animal with a  
XX deficiency in cerebellar granule neurons or their precursors, and may  
XX also be used in promoting mechanoreceptive cell growth and generating  
XX hair cells. The present sequence represents an atonal-associated nucleic  
XX acid sequence referred to in the invention.  
XX Note: The present sequence is not shown in the specification, but  
XX was obtained from GenBank.  
XX  
XX Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;  
XX  
XX Query Match 1.6%; Score 23; DB 22; Length 1412;  
XX Best Local Similarity 100.0%; Pred. No. 0.67;  
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGGCAACCGCATGCACAACCT 754  
Db 445 GAGGCAACCGCATGCACAACCT 467

RESULT 27  
AAF27273  
ID AAF27273 standard; cDNA; 1412 BP.  
XX  
XX AAF27273;  
XX  
XX 24-APR-2001 (first entry)  
XX  
XX Mouse atonal homologue 4A (Math4A) cDNA, SEQ ID NO:37.  
XX  
XX Atonal; homologue; orthologue; atonal-associated protein; deafness;  
XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;  
XX cellular proliferation; cerebellar granule neuron; gene therapy;  
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;  
KW transgenic animal; ss.

```

XX Mus musculus.
OS
XX
XX WO200073764-A2.
XX
XX 07-DEC-2000.
XX
XX 01-JUN-2000; 2000WO-US15410.
XX
XX 01-JUN-1999; 99US-0137060.
XX
XX 19-JAN-2000; 2000US-0176993.
XX
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
XX
XX WPI; 2001-032190/04.
XX
XX P-PSDB; AAB60365.
XX
XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
XX
XX any of its homologs or orthologs, for the treatment of e.g. deafness,
XX
XX osteoarthritis and abnormal cell proliferation -
XX
XX Disclosure; Page -; 142pp; English.
XX
XX The invention relates to the use of atonal-associated nucleic acid or
XX
XX amino acid sequence, or any of its homologues or orthologues as
XX
XX therapeutic agents for the treatment of deafness, partial hearing loss,
XX
XX vestibular effects due to damage or loss of inner hair cells,
XX
XX osteoarthritis and abnormal cell proliferation. The invention also
XX
XX encompasses methods of screening for compounds which affect the
XX
XX expression of an atonal-associated nucleic acid sequence in an animal,
XX
XX and a transgenic animal in which an allele of a native atonal-associated
XX
XX gene is replaced by a heterologous nucleic acid sequence, thus
XX
XX inactivating the atonal-associated allele. The nucleic acids or proteins
XX
XX may be used in a method of treating an animal for hearing impairment,
XX
XX joint disease, balance disorders, abnormal cell proliferation, or other
XX
XX disease related to loss of a functional atonal-associated nucleic acid or
XX
XX protein. They may particularly be used to treat an animal with a
XX
XX deficiency in cerebellar granule neurons or their precursors, and may
XX
XX also be used in promoting mechanoreceptive cell growth and generating
XX
XX hair cells. The present sequence represents an atonal-associated nucleic
XX
XX acid sequence referred to in the invention.
XX
XX Note: The present sequence is not shown in the specification, but
XX
XX was obtained from GenBank.
XX
XX Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;
XX
XX Query Match 1.6%; Score 23; DB 22; Length 1412;
XX
XX Best Local Similarity 100.0%; Pred. No. 0.67;
XX
XX Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 732 GAGCGCAACCGCATGCACACCT 754
XX
XX |||||
XX
XX 445 GAGCGCAACCGCATGCACACCT 467
XX
XX
XX
XX RESULT 28
XX
XX AAD47278
XX
XX ID AAD47278 standard; DNA; 21 BP.
XX
XX
XX AAD47278;
XX
XX AC
XX
XX XX 24-FEB-2003 (first entry)
XX
XX DT
XX
XX XX Human RT-PCR upstream primer for neurogenin-3 DNA isolation.
XX
XX DE
XX
XX XX Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
XX
XX KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX
XX OS Homo sapiens.
XX
XX PA
XX
XX XX WO200274946-A2.
XX
XX PI

```

```

XX
XX 26-SEP-2002.
XX
XX PD
XX
XX XX 26-FEB-2002; 2002WO-DK00130.
XX
XX PE
XX
XX XX 26-FEB-2001; 2001US-271474P.
XX
XX PR
XX
XX XX (NOVO ) NOVO NORDISK AS.
XX
XX PA
XX
XX XX Serup P, Heilmberg H, Gradwohl G;
XX
XX PI
XX
XX WPI; 2003-018804/01.
XX
XX XX
XX
XX Generating insulin-secreting cells from precursor stem cells or adult
XX
XX pancreatic exocrine cells, for generating glucose sensitive insulin
XX
XX secreting beta cells for transplantation, comprises using neurogenin3
XX
XX or NeuroD/beta2 -
XX
XX PT
XX
XX Example 4; Page 28; 66pp; English.
XX
XX PS
XX
XX The invention relates to a method for generating insulin-secreting cells
XX
XX from precursor stem cells or adult pancreatic exocrine cells. The method
XX
XX comprises exposing the precursor cells or exocrine cells to a nucleic
XX
XX acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an
XX
XX activator of ngn3 or NeuroD/beta2 gene expression, under conditions
XX
XX effective to generate the insulin-generating cells from the precursor or
XX
XX exocrine cells. The invention is useful in generating insulin-secreting
XX
XX cells from precursor stem cells or adult pancreatic exocrine cells is
XX
XX useful for generating glucose sensitive insulin secreting beta cells
XX
XX suitable for transplantation, and for in situ development of insulin-
XX
XX secreting cells in a patient. The method is also useful for preventing
XX
XX premature differentiation of precursor stem cells into insulin-secreting
XX
XX beta cells and for identifying compounds that prevent or activate beta
XX
XX cell differentiation. The present sequence is human RT-PCR primer for
XX
XX isolation of neurogenin-3 DNA.
XX
XX CC
XX
XX Sequence 21 BP; 2 A; 7 C; 6 G; 6 T; 0 other;
XX
XX SQ
XX
XX Query Match 1.4%; Score 21; DB 25; Length 21;
XX
XX Best Local Similarity 100.0%; Pred. No. 7.7;
XX
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 460 TGGCGCCTCATCCCTTGATG 480
XX
XX |||||
XX
XX 1 TGGCGCCTCATCCCTTGATG 21
XX
XX
XX
XX RESULT 29
XX
XX AAD47326
XX
XX ID AAD47326 standard; DNA; 21 BP.
XX
XX
XX AAD47326;
XX
XX AC
XX
XX XX 24-FEB-2003 (first entry)
XX
XX DT
XX
XX XX Human RT-PCR forward primer for mouse ngn3 DNA isolation.
XX
XX DE
XX
XX XX Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
XX
XX KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX
XX OS Homo sapiens.
XX
XX PA
XX
XX XX WO200274946-A2.
XX
XX PN
XX
XX XX 26-SEP-2002.
XX
XX PD
XX
XX XX 26-FEB-2002; 2002WO-DK00130.
XX
XX PF
XX
XX XX 26-FEB-2001; 2001US-271474P.
XX
XX PR
XX
XX XX (NOVO ) NOVO NORDISK AS.
XX
XX PA
XX
XX XX Serup P, Heilmberg H, Gradwohl G;
XX
XX PI

```

XX WPI; 2003-018804/01.  
XX  
XX Generating insulin-secreting cells from precursor stem cells or adult  
PT pancreatic exocrine cells, for generating glucose sensitive insulin  
PT secreting beta cells for transplantation, comprises using neurogenin3  
PT or NeuroD/beta2  
XX  
XX Example 5B; Page 37; 66pp; English.  
XX  
XX The invention relates to a method for generating insulin-secreting cells  
CC from precursor stem cells or adult pancreatic exocrine cells. The method  
CC comprises exposing the precursor cells or exocrine cells to: a nucleic  
CC acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an  
CC activator of ngn3 or NeuroD/beta2 gene expression, under conditions  
CC effective to generate the insulin-generating cells from the precursor or  
CC exocrine cells. The invention is useful in generating insulin-secreting  
CC cells from precursor stem cells or adult pancreatic exocrine cells is  
CC useful for generating glucose sensitive insulin secreting beta cells  
CC suitable for transplantation, and for in situ development of insulin-  
CC secreting cells in a patient. The method is also useful for preventing  
CC premature differentiation of precursor stem cells into insulin-secreting  
CC beta cells and for identifying compounds that prevent or activate beta  
CC cell differentiation. The present sequence is human RT-PCR primer for  
CC isolation of mouse ngn3 DNA.  
XX  
XX Sequence 21 BP; 7 A; 7 C; 4 G; 3 T; 0 other;  
SQ  
Query Match 1.4%; Score 21; DB 25; Length 21;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 798 CCGGATGACCGCAACTTACA 818  
Db 1 CCGGATGACCGCAACTTACA 21  
RESULT 30  
AAD46889  
ID AAD46889 standard; DNA; 714 BP.  
XX  
XX AAD46889;  
XX  
XX 27-JAN-2003 (first entry)  
XX  
XX Human neurogenin 1 (Ngn1) gene #2.  
XX  
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;  
KM type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;  
KM islet cell; cell therapy; neurogenin 1; Ngn1; gene; de.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 1..714  
FT CDS /\*tag= a  
FT /product= "Human Ngn1 protein"  
XX  
XX WO200274045-A2.  
XX  
XX 26-SEP-2002.  
XX  
XX 20-MAR-2002; 2002WO-US11166.  
XX  
XX 20-MAR-2001; 2001US-0817360.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX German MS, Lin J;  
XX  
XX WPI; 2002-759853/82.  
XX  
XX P-PSDB; AAE29280.  
XX

PT Producing a mammalian islet cell for treating diabetes mellitus  
PT comprises introducing into a mammalian cell a nucleic acid molecule  
PT encoding neuroendocrine basic helix-loop-helix transcription factor  
XX  
XX  
XX Disclosure; Page 94; 108pp; English.  
XX  
XX The invention relates to a method for producing a mammalian islet cell.  
CC The method comprising introducing into a mammalian cell a nucleic acid  
CC molecule encoding an islet transcription factor for expression of the  
CC islet transcription factor in the cell and for production of islet cell  
CC phenotype in the cell. The islet transcription factor is a neuroendocrine  
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful  
CC for treating type 2 diabetes mellitus and for replacing beta cells lost  
CC to autoimmune destruction in individuals with type 1 diabetes. The method  
CC is useful in cell therapy. The present sequence is human neurogenin 1  
CC (Ngn1) gene.  
XX  
XX Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;  
SQ  
Query Match 1.4%; Score 21; DB 24; Length 714;  
Best Local Similarity 100.0%; Pred. No. 6.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 732 GAGCGCAACCGCATGCAAC 752  
Db 301 GAGCGCAACCGCATGCAAC 321  
RESULT 31  
AAT74891  
ID AAT74891 standard; DNA; 1268 BP.  
XX  
XX AAT74891;  
XX  
XX 02-OCT-1997 (first entry)  
XX  
XX Human neurogenic differentiation protein (NeuroD3) DNA clone 20A1.  
XX  
XX Neurogenic differentiation protein; NeuroD; neuroD3 gene;  
KM transcriptional activator; neuron; pancreas; gastrointestinal;  
KM knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;  
KM ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH 55..768  
FT CDS /\*tag= a  
XX  
XX WO9716548-A1.  
XX  
XX 09-MAY-1997.  
XX  
XX 30-OCT-1996; 96WO-US17532.  
XX  
XX 02-NOV-1995; 95US-0552142.  
XX  
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX  
XX (WEIN/) WEINTRAUB N.  
XX  
XX Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;  
XX  
XX WPI; 1997-272117/24.  
XX  
XX P-PSDB; AAW22440.  
XX  
XX Nucleic acid encoding neurogenic differentiation polypeptide -  
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal  
PT development  
XX  
XX Claim 1; Page 64-65; 81pp; English.  
XX  
XX Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins  
CC (AAW22436-43) from human, mouse and frog have been identified.

CC isolated and sequence. Neurod polypeptides are tissue-specific  
CC basic-helix-loop-helix (bHLH) transcriptional activators involved  
CC in neuronal, endocrine and gastrointestinal development. They were  
CC discovered by expression cloning and screening assays designed to  
CC identify possible bHLH proteins capable of interacting with the  
CC protein product of the Drosophila daughterless gene. Novel neurod2  
CC and neurod3 genes, related to neurod1, have been identified.  
CC Neurod nucleic acids can be used to produce Neurod polypeptides,  
CC construction of test cell lines, as probes, in gene therapy, and to  
CC produce transgenic animals as models of disease.

XX SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;  
XX  
XX Query Match 1.4%; Score 21; DB 18; Length 1268;  
XX Best Local Similarity 100.0%; Pred. No. 6.5;  
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752  
Db 355 GAGCGCAACCGCATGCACAC 375  
|||||  
|||||

RESULT 32  
AAV42932  
ID AAV42932 standard; DNA; 1268 BP.  
XX  
XX AAV42932;  
XX  
XX 25-MAR-2003 (updated)  
DT 21-OCT-1998 (first entry)  
XX  
XX DNA encoding human neurod3 protein, which is a bHLH protein.  
DE  
XX Basic helix-loop-helix; bHLH; neurod; neuroectodermal tumour;  
KM classification; medulloblastoma; human; ds.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT CDS 55..768  
FT /\*tag= a  
FT /product= neurod3

US5795723-A.  
XX  
XX 18-AUG-1998.  
XX  
XX 07-AUG-1997; 97US-0910973.  
XX  
XX 06-MAY-1994; 94US-0239238.  
PR 02-NOV-1995; 95US-0552142.  
PR 30-OCT-1996; 96WO-US17532.  
XX  
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PA  
XX Olson JM, Tapscott SJ;  
PI  
XX WPI: 1998-466661/40.  
DR P-PSDB; AAW71016.  
XX  
XX Classifying neuroectodermal tumours from expression pattern of  
PT basic-helix-loop-helix genes - especially for identifying  
PT medulla:blastoma and assessing its aggressiveness, specifically  
PT associated with expression of bHLH genes neurod 1-3  
XX  
XX Example 11; Columns 61-64; 43pp; English.

XX The present sequence encodes a protein which is a member of the basic  
CC helix-loop-helix (bHLH) protein family, and is designated neurod3. The  
CC bHLH genes are a family of genes associated with vertebrate neuronal,  
CC endocrinal and gastrointestinal development. The observed pattern of  
CC neurod expression distinguishes subclasses of neuroectodermal tumours.  
CC The specification describes a method for the classification of human

CC neuroectodermal tumours. The method comprises measuring, in a tumour  
CC sample, expression of at least one basic bHLH gene and identifying the  
CC tumour subclasses by matching expression to predetermined expression  
CC profiles for known subclasses. For classifying the tumour as a  
CC medulloblastoma, the bHLH gene detected is neurod1 and neurod3.  
CC The method is used to classify neuroectodermal tumours, and to identify  
CC medulloblastoma and for prognosis of this as aggressive.  
CC (updated on 25-MAR-2003 to correct PR field.)  
XX  
XX SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;  
XX  
XX Query Match 1.4%; Score 21; DB 19; Length 1268;  
XX Best Local Similarity 100.0%; Pred. No. 6.5;  
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752  
Db 355 GAGCGCAACCGCATGCACAC 375  
|||||  
|||||

RESULT 33  
ABS56390  
ID ABS56390 standard; DNA; 1268 BP.  
XX  
XX ABS56390;  
XX  
XX 23-JAN-2003 (first entry)  
DT  
XX  
XX Human bHLH family neurod3 genomic DNA, clone 20A1.  
DE  
XX Human; gene; ds; neurod3; neurod; basic-helix-loop-helix; bHLH;  
KM differentiation; neurone; endocrine; gastrointestinal; development;  
KM transgenic; embryo; birth defect; spontaneous abortion; stem cell;  
KM cancer; neural growth factor; tumour; diagnostic; motor; sensory;  
KM traumatic neural injury; hearing; vision; brain; spinal cord;  
KM malabsorption syndrome; gastrointestinal dysmotility syndrome;  
KM Hirsch Prung's disease; therapeutic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX Key Location/Qualifiers  
FT CDS 55..768  
FT /\*tag= a  
FT /product= "Neurod3"  
FT misc\_feature 376..495  
FT /\*tag= b  
FT /note= "HLH coding domain"

US6444463-B1.  
XX  
XX 03-SEP-2002.  
XX  
XX 07-FEB-2000; 2000US-0499227.  
XX  
XX 05-AUG-1998; 98WO-US16417.  
PR 06-MAY-1994; 94US-0239238.  
PR 08-MAY-1995; 95WO-US05741.  
PR 02-NOV-1995; 95US-0552142.  
PR 30-OCT-1996; 96WO-US17532.  
PR 07-AUG-1997; 97US-0910973.  
XX  
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
PA  
XX Tapscott SJ;  
PI  
XX WPI: 2003-056678/05.  
DR P-PSDB; ABG72002.  
XX  
XX New neurogenic differentiation gene, useful in gene therapy to correct  
PT traumatic neural injury that has resulted in loss of motor or sensory  
PT neural function and for constructing recombinant cell lines -  
XX  
XX Claim 1; Column 61-64; 43pp; English.

XX The invention discloses an isolated nucleic acid molecule which encodes a  
 CC functionally active human neurod2 polypeptide. Neurod proteins represent  
 CC a new family within the basic-helix-loop-helix (bHLH) family which are  
 CC implicated in the regulation of differentiation. Neurod proteins are  
 CC particularly involved in neuronal, endocrine and gastrointestinal  
 CC development. The nucleic acid is useful for constructing recombinant cell  
 CC lines, transgenic embryos and animals and for quantifying the level of  
 CC expression of neurod in a cell. Birth defects and spontaneous abortions  
 CC may result from expression of an abnormal neurod protein. The  
 CC polynucleotide sequences permit the establishment of primary cultures of  
 CC proliferating embryonic neuronal stem cells under conditions mimicking  
 CC those that are active in development and cancer. The resultant cell lines  
 CC find use as sources of novel neural growth factors, in assays for  
 CC identifying novel neuronal growth factors which can be used for screening  
 CC anti-cancer drugs capable of driving terminal differentiation in neural  
 CC tumours, for producing antibodies useful in diagnostic assays and for  
 CC screening for compounds capable of modulating the activity of neurod.  
 CC Transformed host cells, nucleic acids and polypeptides are also useful  
 CC for treating sites of traumatic neural injury where motor or sensory  
 CC neural activity has been lost, e.g. hearing or vision loss and brain or  
 CC spinal cord damage. The host cells find use in the treatment of  
 CC malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsch  
 CC Prung's Disease). The cell lines also find use in screening for candidate  
 CC therapeutic agents capable of either substituting for neurod or  
 CC correcting the cellular defect caused by a defective neurod. The sequence  
 CC presented is the human neurod3 genomic DNA, clone 20A1.

SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;

Query Match 1.4%; Score 21; DB 25; Length 1268;  
 Best Local Similarity 100.0%; Pred. No. 6.5;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGGCAACCGCATGCACAC 752  
 Db 355 GAGGCAACCGCATGCACAC 375

RESULT 34

AAAT74890 standard; DNA; 1535 BP.

AC AAT74890;

DT 02-OCT-1997 (first entry)

DE Human neurogenic differentiation protein (Neurod2) DNA clone 14B1.

XX Neurogenic differentiation protein; Neurod; neurod2 gene;

KM transcriptional activator; neuron; pancreas; gastrointestinal;

KM knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;

KM ss.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FT CDS

XX Location/Qualifiers

XX 55..1200

XX /\*tag= a

XX MO9716548-A1.

XX 09-MAY-1997.

XX 30-OCT-1996; 96WO-US17532.

XX 02-NOV-1995; 95US-0552142.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX (WEIN/) WEINTRAUB N.

XX Hollenberg SM, Lee JE, Tapscott SU, Weintraub HM;

XX

DR WPI; 1997-272117/24.  
 DR P-PSDB; AAW22439.  
 XX Nucleic acid encoding neurogenic differentiation polypeptide -  
 PT useful e.g. in regulating neuronal, endocrine and gastrointestinal  
 PT development

PS Claim 1; Page 61-62; 81pp; English.

XX Neurogenic differentiation (Neurod) genes (AAT74887-94) and proteins

CC (AAW22436-43) from human, mouse and frog have been identified

CC isolated and sequenced. Neurod polypeptides are tissue-specific

CC basic-helix-loop-helix (bHLH) transcriptional activators involved

CC in neuronal, endocrine and gastrointestinal development. They were

CC discovered by expression cloning and screening assays designed to

CC identify possible bHLH proteins capable of interacting with the

CC protein product of the Drosophila daughterless gene. Novel neurod2

CC and neurod3 genes, related to neurod1, have been identified.

CC Neurod nucleic acids can be used to produce Neurod polypeptides,

CC construction of test cell lines, as probes, in gene therapy, and to

CC produce transgenic animals as models of disease.

SQ Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;

Query Match 1.4%; Score 21; DB 18; Length 1535;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 729 CGGAGCGCAACCGCATGCAC 749  
 Db 439 CGGAGCGCAACCGCATGCAC 459

RESULT 35

AAV42931 standard; DNA; 1535 BP.

AC AAV42931;

DT 25-MAR-2003 (updated)

DT 21-OCT-1998 (first entry)

DE DNA encoding human neurod2 protein, which is a bHLH protein.

XX Basic helix-loop-helix; bHLH; neurod; neuroectodermal tumour;

KM classification; medulloblastoma; human; ds.

XX Homo sapiens.

OS Homo sapiens.

XX Key

FT CDS

XX Location/Qualifiers

XX 55..1194

XX /\*tag= a

XX /product= neurod2

XX US5795723-A.

XX 18-AUG-1998.

XX 07-AUG-1997; 97US-0910973.

XX 06-MAY-1994; 94US-0239238.

XX 02-NOV-1995; 95US-0552142.

XX 30-OCT-1996; 96WO-US17532.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Olson JM, Tapscott SJ;

XX WPI; 1998-466661/40.

XX P-PSDB; AAW71015.

XX Classifying neuroectodermal tumours from expression pattern of

XX basic-helix-loop-helix genes - especially for identifying

PT medulla:blastoma and assessing its aggressiveness, specifically  
 PT associated with expression of BHLH genes neurod 1-3  
 XX  
 PS Example 11; Columns 57-60; 43pp; English.  
 XX  
 CC The present sequence encodes a protein which is a member of the basic  
 CC helix-loop-helix (bHLH) protein family, and is designated neurod2. The  
 CC bHLH genes are a family of genes associated with vertebrate neuronal,  
 CC endocrinal and gastrointestinal development. The observed pattern of  
 CC neurod expression distinguishes subclasses of neuroectodermal tumours.  
 CC The specification describes a method for the classification of human  
 CC neuroectodermal tumours. The method comprises measuring, in a tumour  
 CC sample, expression of at least one basic bHLH gene and identifying the  
 CC tumour subclasses by matching expression to predetermined expression  
 CC profiles for known subclasses. For classifying the tumour as a  
 CC medulloblastoma, the bHLH gene detected is neurod1 and neurod3.  
 CC The method is used to classify neuroectodermal tumours, and to identify  
 CC medulloblastoma and for prognosis of this as aggressive.  
 CC (Updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SO Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;  
 Query Match 1.4%; Score 21; DB 19; Length 1535;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 729 CGGAGCGCCACCGCATGCAC 749  
 Db 439 CGGAGCGCCACCGCATGCAC 459  
 RESULT 36  
 ABS56389  
 ID ABS56389 standard; DNA; 1535 BP.  
 AC ABS56389;  
 XX  
 DT 23-JAN-2003 (first entry)  
 XX  
 DE Human bHLH family neurod2 genomic DNA, clone 14B1.  
 XX  
 KW Human; gene; ds; neurod3; neurod; basic-helix-loop-helix; bHLH;  
 KW differentiation; neurone; endocrine; gastrointestinal; development;  
 KW transgenic; embryo; birth defect; spontaneous abortion; stem cell;  
 KW cancer; neural growth factor; tumour; diagnostic; motor; sensory;  
 KW traumatic neural injury; hearing; vision; brain; spinal cord;  
 KW malabsorption syndrome; gastrointestinal dysmotility syndrome;  
 KW Hirsch Prung's disease; therapeutic; neurod2.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 55..1194  
 FT /\*tag= a  
 FT /product= "Neurod2"  
 FT misc\_feature 463..582  
 FT /\*tag= b  
 FT /note= "HLH coding domain"  
 FT  
 XX  
 PN US6444463-B1.  
 XX  
 PD 03-SEP-2002.  
 XX  
 PF 07-FEB-2000; 2000US-0499227.  
 XX  
 XX 05-AUG-1998; 98WO-US16417.  
 PR 06-MAY-1994; 94US-0239238.  
 PR 08-MAY-1995; 95MO-US05741.  
 PR 02-NOV-1995; 95US-0552142.  
 PR 30-OCT-1996; 96WO-US17532.  
 PR 07-AUG-1997; 97US-0910973.  
 XX  
 PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX  
 PI Tapscott SJ;  
 XX  
 DR WPI; 2003-056678/05.  
 DR P-PSDB; ABG72001.  
 XX  
 PT New neurogenic differentiation gene, useful in gene therapy to correct  
 PT traumatic neural injury that has resulted in loss of motor or sensory  
 PT neural function and for constructing recombinant cell lines -  
 XX  
 PS Example 11; Column 57-60; 43pp; English.  
 XX  
 CC The invention discloses an isolated nucleic acid molecule which encodes a  
 CC functionally active human neurod3 polypeptide. Neurod proteins represent  
 CC a new family within the basic-helix-loop-helix (bHLH) family which are  
 CC implicated in the regulation of differentiation. Neurod proteins are  
 CC particularly involved in neuronal, endocrine and gastrointestinal  
 CC development. The nucleic acid is useful for constructing recombinant cell  
 CC lines, transgenic embryos and animals and for quantifying the level of  
 CC expression of neurod in a cell. Birth defects and spontaneous abortions  
 CC may result from expression of an abnormal neurod protein. The  
 CC polynucleotide sequences permit the establishment of primary cultures of  
 CC proliferating embryonic neuronal stem cells under conditions mimicking  
 CC those that are active in development and cancer. The resultant cell lines  
 CC find use as sources of novel neural growth factors, in assays for  
 CC identifying novel neuronal growth factors which can be used for screening  
 CC anti-cancer drugs capable of driving terminal differentiation in neural  
 CC tumours, for producing antibodies useful in diagnostic assays and for  
 CC screening for compounds capable of modulating the activity of neurod.  
 CC Transformed host cells, nucleic acids and polypeptides are also useful  
 CC for treating sites of traumatic neural injury where motor or sensory  
 CC neural activity has been lost, e.g. hearing or vision loss and brain or  
 CC spinal cord damage. The host cells find use in the treatment of  
 CC malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsch  
 CC Prung's disease). The cell lines also find use in screening for candidate  
 CC therapeutic agents capable of either substituting for neurod or  
 CC correcting the cellular defect caused by a defective neurod. The sequence  
 CC presented is the human neurod2 genomic DNA, clone 14B1.  
 XX  
 SO Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;  
 Query Match 1.4%; Score 21; DB 25; Length 1535;  
 Best Local Similarity 100.0%; Pred. No. 6.4;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 729 CGGAGCGCCACCGCATGCAC 749  
 Db 439 CGGAGCGCCACCGCATGCAC 459  
 RESULT 37  
 AAA62681  
 ID AAA62681 standard; DNA; 1550 BP.  
 AC AAA62681;  
 XX  
 DT 29-NOV-2000 (first entry)  
 XX  
 DE Human Neurod2 gene.  
 XX  
 KW Human; Neurod2; neurogenic basic helix-loop-helix protein;  
 KW epidermal cell transdifferentiation; gene therapy; cerebroprotective;  
 KW neuroprotective; brain injury; spinal cord injury; stroke;  
 KW neurodegenerative disease; Parkinson's disease; Huntington's disease;  
 KW Alzheimer's disease; neuronal cell generation; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 58..1206  
 FT /\*tag= a  
 FT /product= "Neurod2 protein"  
 FT  
 XX

PN US6087168-A.  
XX  
XX 11-JUL-2000.  
XX  
XX 20-JAN-1999; 99US-0234332.  
XX  
XX 20-JAN-1999; 99US-0234332.  
XX  
XX (CEDA-) CEDARS SINAI MEDICAL CENT.  
XX  
XX Levesque MF, Neuman T;  
XX  
XX WPI; 2000-498200/44.  
XX P-PSDB; AAB14347.  
XX  
XX Converting epidermal cells into neurons, useful for isolating nerve  
PT growth factors or for gene therapy, comprises differentiating cells  
PT and transfecting with vectors with a cDNA coding neurogenic  
PT transcription factors -  
XX  
XX Example 2; Column 19-22; 27pp; English.  
XX  
XX The present sequence is the human neurogenic basic helix-loop-helix  
CC protein (NeuroD2) gene from Genbank. It was used to provide sequence  
CC information for the cloning of NeuroD2 cDNA, which was used to  
CC transfect cultured epidermal cells. This was part of a novel method  
CC for transdifferentiating an epidermal basal cell into a cell having the  
CC morphological, physiological and/or immunological features of a viable  
CC neuronal cell. The method is useful for screening new drugs for  
CC treating a nervous system disorder, or for isolating a novel nerve  
CC growth factor. The transdifferentiated cell is useful in both cell and  
CC gene therapies aimed at alleviating various neurological disorders. The  
CC cell or gene therapy approach involves the use of autologous  
CC transplantation or grafting of the newly created neuronal cells as  
CC treatment for brain or spinal cord injury, stroke and neurodegenerative  
CC diseases (e.g. Parkinson's disease, Huntington's disease or Alzheimer's  
CC disease).  
XX  
XX Sequence 1550 BP; 250 A; 569 C; 486 G; 243 T; 2 other;  
SQ  
XX  
XX Query Match 1.4%; Score 21; DB 21; Length 1550;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 729 CGGAGCGCAACCGCATGCAC 749  
Db 442 CGGAGCGCAACCGCATGCAC 462  
RESULT 38  
AAD46888  
ID AAD46888 standard; DNA; 1665 BP.  
XX  
XX AAD46888;  
XX  
XX 27-JAN-2003 (first entry)  
XX  
XX Human neurogenin 1 (Ngn1) gene #1.  
XX  
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;  
XX type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;  
XX islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX CDS 260..973  
XX /tag= a  
XX /product= "Human Ngn1 protein"  
XX  
XX MO200274045-A2.  
XX  
XX 26-SEP-2002.

XX  
XX 20-MAR-2002; 2002WO-US11166.  
XX  
XX 20-MAR-2001; 2001US-0817360.  
XX  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX German MS, Lin J;  
XX  
XX WPI; 2002-759853/82.  
XX P-PSDB; AAE29279.  
XX  
XX Producing a mammalian islet cell for treating diabetes mellitus  
PT comprises introducing into a mammalian cell a nucleic acid molecule  
PT encoding neuroendocrine basic helix-loop-helix transcription factor -  
XX  
XX Disclosure; Page 93; 108pp; English.  
XX  
XX The invention relates to a method for producing a mammalian islet cell.  
CC The method comprising introducing into a mammalian cell a nucleic acid  
CC molecule encoding an islet transcription factor for expression of the  
CC islet transcription factor in the cell and for production of islet cell  
CC phenotype in the cell. The islet transcription factor is a neuroendocrine  
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful  
CC for treating type 2 diabetes mellitus and for replacing beta cells lost  
CC to autoimmune destruction in individuals with type 1 diabetes. The method  
CC is useful in cell therapy. The present sequence is human neurogenin 1  
CC (Ngn1) gene.  
XX  
XX Sequence 1665 BP; 344 A; 561 C; 443 G; 317 T; 0 other;  
SQ  
XX  
XX Query Match 1.4%; Score 21; DB 24; Length 1665;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 732 GAGCGCAACCGCATGCACAC 752  
Db 560 GAGCGCAACCGCATGCACAC 580  
RESULT 39  
AAL04043/c  
ID AAL04043 standard; DNA; 2776 BP.  
XX  
XX AAL04043;  
XX  
XX 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 6731.  
XX  
XX Human; reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX MO200155320-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01339.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.

[illegible]

PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249269.
PR	17-NOV-2000;	2000US-0249279.
PR	17-NOV-2000;	2000US-0249289.
PR	01-DEC-2000;	2000US-0250300.
PR	01-DEC-2000;	2000US-0250310.
PR	01-DEC-2000;	2000US-0250311.
PR	01-DEC-2000;	2000US-0250312.
PR	01-DEC-2000;	2000US-0250313.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251989.
PR	06-DEC-2000;	2000US-0251719.
PR	06-DEC-2000;	2000US-0251719.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-465570/50.	
DR		
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen	
XX	is used in preventing, treating or ameliorating a medical condition -	
PT		
XX		
PS	Disclosure; SEQ ID NO 6731; 1297bp + Sequence Listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a	
CC	number of human reproductive system related antigens. These can be used	
CC	in the prevention and treatment of reproductive system disorders,	
CC	including cancer. The present sequence is a genomic sequence encoding a	
CC	protein of the invention.	
XX		
SQ	Sequence 2776 BP; 631 A; 851 C; 749 G; 545 T; 0 other:	
Query Match	1.4%; Score 21; DB 22; Length 2776;	
Best Local Similarity	100.0%; Pred. No. 6.3;	



```
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1112 AACAGCCCTGGCGGTGGGC 1132
Db 938 AACAGCCCTGGCGGTGGGC 918

RESULT 40
AAL04045/C
ID AAL04045 standard; DNA; 2776 BP.
AC AAL04045;
XX
XX 21-NOV-2001 (first entry)
DT
XX Human reproductive system related antigen DNA SEQ ID NO: 6733.
DE Human reproductive system related antigen; reproductive system disorder;
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX WO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225577.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226686.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
```

```
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
```

XX	PA	(HUMA-) HUMAN GENOME SCI INC.
XX	PI	Rosen CA, Barash SC, Ruben SM;
XX	DR	WPI; 2001-465570/50.
XX	PT	
XX	PT	Isolated nucleic acid molecule encoding a reproductive system antigen -
XX	PT	is used in preventing, treating or ameliorating a medical condition -
PS	PS	Disclosure; SEQ ID NO 6733; 1297pp + Sequence Listing; English.
CC	CC	The present invention provides the protein and coding sequences of a
CC	CC	number of human reproductive system related antigens. These can be used
CC	CC	in the prevention and treatment of reproductive system disorders,
CC	CC	including cancer. The present sequence is a genomic sequence encoding a
CC	CC	protein of the invention.
XX	XX	
XX	XX	Sequence 2776 BP; 631 A; 853 C; 747 G; 545 T; 0 other;

Query Match	1.4%;	Score 21;	DB 22;	Length 2776;
Best Local Similarity	100.0%;	Pred. No. 6.3;		
Matches 21;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

**OY**      1112 AACAGCCCTGGGCGGTGGC 1132  
         |||||  
**Db**      938 AACAGCCCTGGGCGGTGGC 918

## RESULT 41

AAK68475

AC AAK68475;

DT 06-NOV-2001 (first entry)

DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23287.
DE	

KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN W0200157182-A2.

PD 09-AUG-2001

PF 17-JAN-2001; 2001WO-US01354.

PR 31-JAN-2000; 2000US-0179065.

PR 24-FEB-2000; 2000US-0184664.

PR 16-MAR-2000; 2000US-0189874.

PR	17-MAR-2000	2000US-01090076
PR	18-MAR-2000	2000US-01091233
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209467
PR	28-JUN-2000	2000US-02144866
PR	30-JUN-2000	2000US-02151305
PR	07-JUL-2000	2000US-02166347
PR	07-JUL-2000	2000US-0216680
PR	11-JUL-2000	2000US-0217487
PR	11-JUL-2000	2000US-0217496
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-0220963
PR	26-JUL-2000	2000US-0220964
PR	14-AUG-2000	2000US-0224518
PR	14-AUG-2000	2000US-0224519
PR	14-AUG-2000	2000US-0225513
PR	14-AUG-2000	2000US-0225514
PR	14-AUG-2000	2000US-0225666
PR	14-AUG-2000	2000US-0225677
PR	14-AUG-2000	2000US-0225447
PR	14-AUG-2000	2000US-02254570
PR	14-AUG-2000	2000US-0225470
PR	14-AUG-2000	2000US-0225757
PR	14-AUG-2000	2000US-0225758
PR	14-AUG-2000	2000US-0225759
PR	18-AUG-2000	2000US-0226579
PR	22-AUG-2000	2000US-0226681
PR	22-AUG-2000	2000US-0226686
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227009
PR	30-AUG-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229346
PR	05-SEP-2000	2000US-0229509
PR	06-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231142
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231143
PR	08-SEP-2000	2000US-0231144
PR	08-SEP-2000	2000US-0231143
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233064
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234223
PR	21-SEP-2000	2000US-0234223
PR	25-SEP-2000	2000US-0234977
PR	25-SEP-2000	2000US-0234978
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236127
PR	29-SEP-2000	2000US-0236167
PR	29-SEP-2000	2000US-0236168
PR	29-SEP-2000	2000US-0236169
PR	29-SEP-2000	2000US-0236170
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239935

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-02559678.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX  
XX PT useful for preventing, diagnosing and/or treating cancers and  
XX  
XX PT metastasis -  
XX  
XX PS Disclosure; SEQ ID NO 23287; 3071bp + Sequence Listing; English.  
XX  
XX CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
XX  
XX CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic

CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK67694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
XX SQ Sequence 2776 BP; 545 A; 749 C; 851 G; 631 T; 0 other;  
XX  
XX Query Match 1.4%; Score 21; DB 22; Length 2776;  
XX Best Local Similarity 100.0%; Pred. No. 6.3;  
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1112 AACAGGCCCTGGCGGTGGGC 1132  
DB 1839 AACAGGCCCTGGCGGTGGGC 1859  
RESULT 42  
AAK6476  
ID AAK6476 standard; DNA; 2776 BP.  
XX  
XX AC AAK6476;  
XX  
XX DT 06-NOV-2001 (first entry)  
XX  
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23288.  
XX  
XX DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX  
XX KW cystostatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200157182-A2.  
XX  
XX PD 09-AUG-2001.  
XX  
XX PF 17-JAN-2001; 2001WO-US01354.  
XX  
XX PR 31-JAN-2000; 2000US-0179065.  
XX PR 04-FEB-2000; 2000US-0180628.  
XX PR 24-FEB-2000; 2000US-0184664.  
XX PR 02-MAR-2000; 2000US-0186350.  
XX PR 16-MAR-2000; 2000US-0189874.  
XX PR 17-MAR-2000; 2000US-0190076.  
XX PR 18-APR-2000; 2000US-0198123.  
XX PR 19-MAY-2000; 2000US-0205515.  
XX PR 07-JUN-2000; 2000US-0209467.  
XX PR 28-JUN-2000; 2000US-0214886.  
XX PR 30-JUN-2000; 2000US-0215135.  
XX PR 07-JUL-2000; 2000US-0216647.  
XX PR 07-JUL-2000; 2000US-0216880.  
XX PR 11-JUL-2000; 2000US-0217487.  
XX PR 11-JUL-2000; 2000US-0217496.  
XX PR 14-JUL-2000; 2000US-0218290.  
XX PR 26-JUL-2000; 2000US-0220963.  
XX PR 14-AUG-2000; 2000US-0220964.  
XX PR 14-AUG-2000; 2000US-0224518.  
XX PR 14-AUG-2000; 2000US-0224519.  
XX PR 14-AUG-2000; 2000US-0225213.  
XX PR 14-AUG-2000; 2000US-0225214.  
XX PR 14-AUG-2000; 2000US-0225266.  
XX PR 14-AUG-2000; 2000US-0225267.  
XX PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225557.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 12-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249221.  
 PR 17-NOV-2000; 2000US-0249224.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249246.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251859.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254979.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-483426/52.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 PS  
 XX  
 PS Disclosure; SEQ ID NO 23288; 3071pp + Sequence Listing; English.  
 CC  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK61924 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
 CC represent sequences used in the exemplification of the present invention.  
 CC  
 XX  
 SQ Sequence 2776 BP; 545 A; 747 C; 853 G; 631 T; 0 other;  
 Query Match 1.4%; Score 21; DB 22; Length 2776;  
 Best Local Similarity 100.0%; Pred. No. 6.3;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGCCCTGGCGGTGGGC 1132  
 |||||  
 Db 1839 AACAGCCCTGGCGGTGGGC 1859

RESULT 43  
 ABS69045/C  
 ID ABS69045 standard; DNA; 352 BP.  
 XX  
 AC ABS69045;  
 XX  
 DT 21-NOV-2002 (first entry)  
 XX

Novel murine polynucleotide isolated using gene trap technology #108.

Mouse; gene trapped sequence; GTS; functional genomic analysis;  
 phage display system; gene chip; temporal gene expression;  
 tissue specific gene expression; antisense inhibition; gene targeting;  
 development disorder; cell differentiation disorder; aging; cancer;  
 autoimmune disease; lupus; inflammatory disorder; skin disorder;  
 degenerative disorder; ds.

Mus musculus.

US2002102543-A1.

01-AUG-2002.

30-NOV-2000; 2000US-0728445.

01-DEC-1999; 99US-168358P.

(FRIE/) FRIEDRICH G.  
 (ZAMB/) ZAMBROWICZ B.  
 (SAND/) SANDS A T.

Friedrich G, Zambrowicz B, Sands AT;

WPI; 2002-690598/74.

Novel murine polynucleotides that individually identify novel genes  
 into which a retroviral gene trap vector has integrated, useful in  
 genomic analysis and in discovery, development of therapeutic and  
 diagnostic agents -

Claim 1; Page 57; 296pp; English.

The invention describes an isolated murine polynucleotide (I) comprising  
 a contiguous stretch of at least 60 nucleotides of one of 265-677  
 nucleotide 891 OMNIBANK gene trapped sequences (GTS) (S), given in the  
 specification. The novel genes and cells are useful in functional  
 genomic analysis and in the discovery and development of new therapeutic  
 and diagnostic agents and methods. (I) is useful for identifying the  
 coding regions of the murine genome, to isolate cDNAs, genomic clones,  
 or full-length genes/polynucleotides or homologues, heterologues,  
 paralogues, or orthologues that are capable of hybridizing to one or more  
 of the GTS under stringent conditions. (I) can be incorporated into a  
 phage display system that can be used to screen for proteins, or other  
 ligands, that are capable of binding an amino acid sequence encoded by  
 an oligonucleotide or polynucleotide sequence in at least one of the TS  
 sequences. (I) is useful in addressable arrays, such as gene chips, to  
 identify and characterize temporal and tissue specific gene expression,  
 to identify the gene of interest from many sources and for genetic  
 manipulations such as antisense inhibition and gene targeting. Decreasing  
 the level of expression of (I) and/or down regulating the activity of  
 peptides or proteins encoded by (I) is useful for treating development  
 and cell differentiation disorders, aging, cancer, autoimmune disease,  
 lupus, inflammatory disorders, skin disorders and degenerative  
 disorders. This sequence represents a murine cDNA isolated using gene  
 trap technology.

Sequence 352 BP; 97 A; 89 C; 84 G; 82 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 352;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1010 CACAGCCTCATTTGAGAGGT 1029  
 |||||  
 Db 209 CACAGCCTCATTTGAGAGGT 190

RESULT 44  
 ABQ49522/C  
 ID ABQ49522 standard; DNA; 592 BP.  
 XX  
 AC ABQ49522;  
 XX  
 DT 12-JUL-2002 (first entry)  
 XX

Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;  
 drug; side effect; cancer; central nervous system; cardiovascular;  
 gastrointestinal; respiratory system; single nucleotide polymorphism;  
 SNP; cell differentiation; ds.

Homo sapiens.

WO200218632-A2.

07-MAR-2002.

01-SEP-2001; 2001WO-EP10074.

01-SEP-2000; 2000DE-1043826.

05-SEP-2000; 2000DE-1044543.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K, Guecig D;

WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful  
 for diagnosis and prognosis, comprises selective hybridization of  
 amplicons from chemically treated DNA -

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of  
 methylation of a particular cytosine in a motif 5'-CpG-3', present in a  
 genomic sample of DNA. The sample is treated chemically to convert  
 cytosine (C) but not methylated C, to uracil, then part of the genomic  
 DNA that contains the target C is amplified to form a labeled amplicon.  
 The amplicon is hybridized to two classes, each with at least one  
 member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers  
 and the degree of hybridization to both classes is determined from the  
 label on the amplicon. From the ratio of labels hybridized to the two  
 classes of oligomers, the degree of methylation is calculated. The method  
 is used: (i) for diagnosis and/or prognosis of side effects of  
 therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders  
 of the central nervous, cardiovascular, gastrointestinal and respiratory  
 systems etc., particularly by detecting mutations or single nucleotide  
 polymorphisms (SNP's); and (ii) for differentiation of cell or tissue  
 types and for investigating cell differentiation. The method allows the  
 methylation status of many C residues to be determined simultaneously.  
 ABQ13410-ABQ94121 represent genomic DNA sequences used to illustrate the  
 method for determining the degree of cytosine methylation described in  
 the disclosure of the invention.

Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 592;

Best Local Similarity 100.0%; Pred. No. 21;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 834 CGCTTCGCCCACTACAT 853  
DB 68 CGCTTCGCCCACTACAT 49

## RESULT 45

ABQ49523  
ID ABQ49523 standard; DNA, 592 BP.

XX ABQ49523;

XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

KW SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-BP10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A. Piepenbrock C, Berlin K, Guetig D;

XX WPI: 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of

PT amplicons from chemically treated DNA -

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

XX cytosine (C) but not methylated C, to uracil, then part of the genomic

XX DNA that contains the target C is amplified to form a labeled amplicon.

XX The amplicon is hybridised to two classes, each with at least one

XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers

XX and the degree of hybridisation to both classes is determined from the

XX label on the amplicon. From the ratio of labels hybridised to the two

XX classes of oligomers, the degree of methylation is calculated. The method

XX is used: (i) for diagnosis and/or prognosis of side effects of

XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

XX of the central nervous, cardiovascular, gastrointestinal and respiratory

XX systems etc., particularly by detecting mutations or single nucleotide

XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue

XX types and for investigating cell differentiation. The method allows the

XX methylation status of many C residues to be determined simultaneously.

XX ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the

XX method for determining the degree of cytosine methylation described in

XX the disclosure of the invention.

XX Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

XX Query Match 1.4%; Score 20; DB 24; Length 592;

XX Best Local Similarity 100.0%; Pred. No. 21;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 834 CGCTTCGCCCACTACAT 853  
DB 525 CGCTTCGCCCACTACAT 544

Search completed: January 26, 2004, 22:22:39  
Job time : 428 secs



## RESULT 2

US-10-004-717-24  
; Sequence 24, Application US/10004717  
; Publication No. US2002019265A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGHEI, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
; FILE REFERENCE: P01899US4  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 24  
; LENGTH: 861  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-004-717-24

Query Match 6.3%; Score 92; DB 14; Length 861;  
Best Local Similarity 100.0%; Pred. No. 3.8e-37;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GGGCTGATGCGCTGGCGGCTGTCTGCGCCACTTCCCGAGTACCACTTACAAG 821  
DB 463 GGGCTGATGCGCTGGCGGCTGTCTGCGCCACTTCCCGAGTACCACTTACAAG 522  
OY 822 ATCGAGACCTGCGCTTCCGCCCACTACAT 853  
DB 523 ATCGAGACCTGCGCTTCCGCCCACTACAT 554

## RESULT 3

US-09-817-360-3  
; Sequence 3, Application US/09817360  
; Patent No. US2002001566A1  
; GENERAL INFORMATION:  
; APPLICANT: German, Michael S.  
; APPLICANT: Lin, Joseph  
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
; FILE REFERENCE: UCSF-129CIP  
; CURRENT APPLICATION NUMBER: US/09/817,360  
; CURRENT FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 09/535,145  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 60/128,180  
; PRIOR FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1861  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-817-360-3

Query Match 6.3%; Score 92; DB 9; Length 1861;  
Best Local Similarity 100.0%; Pred. No. 3.5e-37;  
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GCGCTGATGCGCTGGCGGCTGTCTGCGCCACTTCCCGAGTACCACTTACAAG 821  
DB 1396 GCGCTGATGCGCTGGCGGCTGTCTGCGCCACTTCCCGAGTACCACTTACAAG 1455

OY 822 ATCGAGACCTGCGCTTCCGCCCACTACAT 853  
DB 1456 ATCGAGACCTGCGCTTCCGCCCACTACAT 1487

## RESULT 4

US-09-908-975-4140  
; Sequence 4140, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, AVI  
; APPLICANT: WASSERMAN, ALON  
; APPLICANT: MINTZ, ELI  
; APPLICANT: MINTZ, LIAT  
; APPLICANT: FAIGLER, SIMCHON  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4140  
; LENGTH: 65  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-09-908-975-4140

Query Match 2.5%; Score 36; DB 13; Length 65;  
Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1281 ATTGAGGCTGATCTCTTAACTCTCTAGTGT 1316  
DB 30 ATTGAGGCTGATCTCTTAACTCTCTAGTGT 65

## RESULT 5

US-09-908-975-30269  
; Sequence 30269, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, AVI  
; APPLICANT: WASSERMAN, ALON  
; APPLICANT: MINTZ, ELI  
; APPLICANT: MINTZ, LIAT  
; APPLICANT: FAIGLER, SIMCHON  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 30269  
; LENGTH: 65  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-908-975-30269

Query Match 2.4%; Score 35; DB 13; Length 65;  
Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Oy 547 TCACTTCATTCACCCGACTAGCCACTCTC 581  
Db 1 TCAGTTCCATTCACCCGACTAGCCACTCTC 35

## RESULT 6

US-09-817-360-1  
; Sequence 1, Application US/09817360  
; Patent No. US20020015696A1  
; GENERAL INFORMATION:  
; APPLICANT: German, Michael S.  
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS  
; TITLE OF INVENTION: AND DELIVERY OF INSULIN  
; FILE REFERENCE: UCSF-129CIP  
; CURRENT APPLICATION NUMBER: US/09/817,360  
; PRIOR FILING DATE: 2001-03-20  
; PRIOR APPLICATION NUMBER: 09/535,145  
; PRIOR FILING DATE: 2000-03-24  
; PRIOR APPLICATION NUMBER: 60/128,180  
; PRIOR FILING DATE: 1999-04-06  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 1  
; LENGTH: 5340  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-09-817-360-1

Query Match 2.2%; Score 32; DB 9; Length 5340;  
Best Local Similarity 100.0%; Pred. No. 4.5e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 672 AAGAGCGAGTTGGCAGCTGAGCAGAGGAGG 703  
Db 3235 AAGAGCGAGTTGGCAGCTGAGCAGAGGAGG 3266

## RESULT 7

US-09-992-665-221/c  
; Sequence 221, Application US/09992665  
; Publication No. US20030092009A1  
; GENERAL INFORMATION:  
; APPLICANT: Kaia Palm  
; TITLE OF INVENTION: PROFILING TUMOR SPECIFIC MARKERS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF NEOPLASTIC DISEASE  
; FILE REFERENCE: CEMINES.002A  
; CURRENT APPLICATION NUMBER: US/09/992,665  
; CURRENT FILING DATE: 2001-11-13  
; PRIOR APPLICATION NUMBER: 60/249,508  
; PRIOR FILING DATE: 2000-11-16  
; NUMBER OF SEQ ID NOS: 380  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO: 221  
; LENGTH: 26  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE: OTHER INFORMATION: Probe  
US-09-992-665-221

Query Match 1.8%; Score 26; DB 11; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.01;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 673 AGAGCGAGTTGGCAGCTGAGCAGAGCAG 698  
Db 26 AGAGCGAGTTGGCAGCTGAGCAGAGCAG 1

RESULT 8  
US-08-722-570-13  
; Sequence 13, Application US/08722570

; Publication No. US20030044887A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David J.

; APPLICANT: Ma, Qifu

; TITLE OF INVENTION: NEUROGENIN

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESS: Flehr, Hobbach, Teat, Albritton & Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/722,570

; FILING DATE: 27-SEP-1996

; CLASSIFICATION: 5365

; ATTORNEY/AGENT INFORMATION:

; NAME: Silva, Robin M.

; REGISTRATION NUMBER: 38,304

; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 738 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA

US-08-722-570-13

Query Match 1.6%; Score 23; DB 8; Length 738;  
Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 732 GAGCGCAACCGCAGCAGCAACT 754  
Db 304 GAGCGCAACCGCAGCAGCAACT 326

## RESULT 9

US-10-004-717-20  
; Sequence 20, Application US/10004717  
; Publication No. US20020192665A1  
; GENERAL INFORMATION:  
; APPLICANT: ZOGHBI, HUDA Y.  
; APPLICANT: YANG, QI  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
; TITLE OF INVENTION: ACTUAL ASSOCIATED SEQUENCE FOR DEPRESS, AND ABNORMAL CELL PROLIFERATION  
; FILE REFERENCE: P01899054  
; CURRENT APPLICATION NUMBER: US/10/004,717  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: 09/585,645  
; PRIOR FILING DATE: 2000-06-01  
; PRIOR APPLICATION NUMBER: 60/176,993  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: 60/137,060  
; PRIOR FILING DATE: 1999-06-01  
; NUMBER OF SEQ ID NOS: 69  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 20  
; LENGTH: 790  
; TYPE: DNA  
; ORGANISM: chicken

US-10-004-717-20

Query Match 1.6%; Score 23; DB 14; Length 790;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
DB 374 GAGCGCAACCGCATGCACACCT 396

RESULT 10  
US-10-004-717-30  
Sequence 30, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899054  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 30  
LENGTH: 1385  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-30

Query Match 1.6%; Score 23; DB 14; Length 1385;  
Best Local Similarity 100.0%; Pred. No. 0.25;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
DB 742 GAGCGCAACCGCATGCACACCT 764

RESULT 11  
US-10-004-717-6  
Sequence 6, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899054  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 1412  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-6

Query Match 1.6%; Score 23; DB 14; Length 1412;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
DB 445 GAGCGCAACCGCATGCACACCT 467

RESULT 12  
US-10-004-717-37  
Sequence 37, Application US/10004717  
Publication No. US2002019265A1  
GENERAL INFORMATION:  
APPLICANT: ZOGHBI, HUDA Y.  
APPLICANT: YANG, QI  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN  
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,  
TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION  
FILE REFERENCE: P01899054  
CURRENT APPLICATION NUMBER: US/10/004,717  
CURRENT FILING DATE: 2002-08-16  
PRIOR APPLICATION NUMBER: 09/585,645  
PRIOR FILING DATE: 2000-06-01  
PRIOR APPLICATION NUMBER: 60/176,993  
PRIOR FILING DATE: 2000-01-19  
PRIOR APPLICATION NUMBER: 60/137,060  
PRIOR FILING DATE: 1999-06-01  
NUMBER OF SEQ ID NOS: 69  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 37  
LENGTH: 1412  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-004-717-37

Query Match 1.6%; Score 23; DB 14; Length 1412;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754  
DB 445 GAGCGCAACCGCATGCACACCT 467

RESULT 13  
US-10-090-011-1  
Sequence 1, Application US/10090011  
Publication No. US20030082810A1  
GENERAL INFORMATION:  
APPLICANT: Serup, Palle  
APPLICANT: Heimberg, Harry  
APPLICANT: Gradwohl, Gerard  
TITLE OF INVENTION: Methods For Generating Insulin-Secreting  
TITLE OF INVENTION: Cells Suitable for Transplantation  
FILE REFERENCE: 6246.200-US  
CURRENT APPLICATION NUMBER: US/10/090,011  
CURRENT FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: US 60/271,474  
PRIOR FILING DATE: 2001-02-26  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 21  
TYPE: DNA  
ORGANISM: Homo Sapien  
US-10-090-011-1

Query Match 1.4%; Score 21; DB 15; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 460 TGGGGCCTCATCCCTTGATG 480  
DB 1 TGGGGCCTCATCCCTTGATG 21

DB 29 AACAGGCCCTGGGCGGTGGGC 49  
Search completed: January 29, 2004, 22:30:14  
Job time : 568 secs

RESULT 14  
US-10-090-011-49  
; Sequence 49, Application US/10090011  
; Publication No. US20030082810A1  
; GENERAL INFORMATION:  
; APPLICANT: Serup, Palle  
; APPLICANT: Heimberg, Harry  
; APPLICANT: Gradwohl, Gerard  
; TITLE OF INVENTION: Methods For Generating Insulin-Secreting  
; TITLE OF INVENTION: Cells Suitable for Transplantation  
; FILE REFERENCE: 6246.200-US  
; CURRENT APPLICATION NUMBER: US/10/090,011  
; CURRENT FILING DATE: 2002-02-26  
; PRIOR APPLICATION NUMBER: US 60/271,474  
; PRIOR FILING DATE: 2001-02-26  
; NUMBER OF SEQ ID NOS: 70  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Homo Sapien  
US-10-090-011-49

Query Match 1.4%; Score 21; DB 15; Length 21;  
Best Local Similarity 100.0%; Pred. No. 4.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 798 CCGGATGACGCCAAGCTTACA 818  
DB 1 CCGGATGACGCCAAGCTTACA 21

RESULT 15  
US-10-029-386-8222  
; Sequence 8222, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DRIVEN SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 8222  
; LENGTH: 576  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
FEATURE:  
; OTHER INFORMATION: MAP TO CHR16.3  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 67  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 64  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.8  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 41  
; OTHER INFORMATION: EST\_HUMAN HIT: BG166801.1, EVALUE 0.00e+00  
; OTHER INFORMATION: NT HIT: g14779902, EVALUE 3.00e-95  
; OTHER INFORMATION: SWISSPROT HIT: P25444, EVALUE 8.00e-58  
US-10-029-386-8222

Query Match 1.4%; Score 21; DB 13; Length 576;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1112 AACAGGCCCTGGGCGGTGGGC 1132

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:42:42 ; Search time 2899 Seconds  
(without alignments)  
12240.267 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460  
Sequence: 1 gcacgtacgcagagagagcagc.....agagtcacccaacccagctgt 1460

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrt1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	92	6.3	593	28	AZ296526
2	60	4.1	1025	14	BY708009
3	60	4.1	1540	11	AK008017
4	24	1.6	923	13	BX445903

5	23	1.6	504	4	BX524720	BX524720 RZPD Mus
6	23	1.6	517	13 <td>BU053975</td> <td>BU053975 UI-M-PD0-</td>	BU053975	BU053975 UI-M-PD0-
7	23	1.6	600	13 <td>BG808248</td> <td>BG808248 2083-52 M</td>	BG808248	BG808248 2083-52 M
8	23	1.6	600	13 <td>BU924937</td> <td>BU924937 7103-91 M</td>	BU924937	BU924937 7103-91 M
9	23	1.6	673	13 <td>BU709606</td> <td>BU709606 UI-M-PRO-</td>	BU709606	BU709606 UI-M-PRO-
10	23	1.6	687	14 <td>CA945402</td> <td>CA945402 UI-M-PD0-</td>	CA945402	CA945402 UI-M-PD0-
11	23	1.6	689	13 <td>BU058877</td> <td>BU058877 UI-M-PRO-</td>	BU058877	BU058877 UI-M-PRO-
12	23	1.6	730	13 <td>BU612495</td> <td>BU612495 UI-M-PRO-</td>	BU612495	BU612495 UI-M-PRO-
13	23	1.6	814	13 <td>BQ178789</td> <td>BQ178789 UI-M-EV0-</td>	BQ178789	BQ178789 UI-M-EV0-
14	23	1.6	823	13 <td>BU054481</td> <td>BU054481 UI-M-PD0-</td>	BU054481	BU054481 UI-M-PD0-
15	23	1.6	932	14 <td>CA979119</td> <td>CA979119 AGENCOURT</td>	CA979119	CA979119 AGENCOURT
16	23	1.6	1001	9 <td>AL540071</td> <td>AL540071 AL540071</td>	AL540071	AL540071 AL540071
17	23	1.6	1269	12 <td>BG854922</td> <td>BG854922 1024041C0</td>	BG854922	BG854922 1024041C0
18	22	1.5	589	13 <td>BU775885</td> <td>BU775885 SGBRD03</td>	BU775885	BU775885 SGBRD03
19	21	1.4	319	10 <td>BE936551</td> <td>BE936551 RCS-NT005</td>	BE936551	BE936551 RCS-NT005
20	21	1.4	464	28 <td>AQ753599</td> <td>AQ753599 HS 2117 A</td>	AQ753599	AQ753599 HS 2117 A
21	21	1.4	536	28 <td>AZ506899</td> <td>AZ506899 IMQ348G13</td>	AZ506899	AZ506899 IMQ348G13
22	21	1.4	704	13 <td>BU057851</td> <td>BU057851 UI-M-PRO-</td>	BU057851	BU057851 UI-M-PRO-
23	21	1.4	718	10 <td>BE263765</td> <td>BE263765 601194122</td>	BE263765	BE263765 601194122
24	21	1.4	740	14 <td>CA319439</td> <td>CA319439 UI-M-FW0-</td>	CA319439	CA319439 UI-M-FW0-
25	21	1.4	829	13 <td>BU611678</td> <td>BU611678 UI-M-FW0-</td>	BU611678	BU611678 UI-M-FW0-
26	21	1.4	829	12 <td>BI910102</td> <td>BI910102 603067946</td>	BI910102	BI910102 603067946
27	21	1.4	835	14 <td>CA320553</td> <td>CA320553 UI-M-FW0-</td>	CA320553	CA320553 UI-M-FW0-
28	21	1.4	913	10 <td>BE204175</td> <td>BE204175 601867625</td>	BE204175	BE204175 601867625
29	21	1.4	953	9 <td>AU067624</td> <td>AU067624 AU067624</td>	AU067624	AU067624 AU067624
30	21	1.4	962	13 <td>BQ686909</td> <td>BQ686909 AGENCOURT</td>	BQ686909	BQ686909 AGENCOURT
31	21	1.4	1022	12 <td>BM563662</td> <td>BM563662 AGENCOURT</td>	BM563662	BM563662 AGENCOURT
32	21	1.4	1039	10 <td>BE780690</td> <td>BE780690 601469349</td>	BE780690	BE780690 601469349
33	21	1.4	1042	10 <td>BG419220</td> <td>BG419220 602445870</td>	BG419220	BG419220 602445870
34	20	1.4	241	9 <td>AA113743</td> <td>AA113743 WSE Pyroc</td>	AA113743	AA113743 WSE Pyroc
35	20	1.4	351	14 <td>T70947</td> <td>T70947 yc49c05.r1</td>	T70947	T70947 yc49c05.r1
36	20	1.4	397	12 <td>BI445873</td> <td>BI445873 da133c04</td>	BI445873	BI445873 da133c04
37	20	1.4	444	9 <td>AV944801</td> <td>AV944801 AV944801</td>	AV944801	AV944801 AV944801
38	20	1.4	476	12 <td>BM253695</td> <td>BM253695 514849 MA</td>	BM253695	BM253695 514849 MA
39	20	1.4	550	9 <td>AV939670</td> <td>AV939670 AV939670</td>	AV939670	AV939670 AV939670
40	20	1.4	909	14 <td>CD325295</td> <td>CD325295 AGENCOURT</td>	CD325295	CD325295 AGENCOURT
41	20	1.4	971	29 <td>CNS01YSD</td> <td>AL173398 Tetradon</td>	CNS01YSD	AL173398 Tetradon
42	20	1.4	971	29 <td>CNS02YSD</td> <td>AL219506 Tetradon</td>	CNS02YSD	AL219506 Tetradon
43	20	1.4	1010	13 <td>BX377288</td> <td>BX377288 BX377288</td>	BX377288	BX377288 BX377288
44	20	1.4	1075	12 <td>BM810917</td> <td>BM810917 AGENCOURT</td>	BM810917	BM810917 AGENCOURT
45	19	1.3	201	10 <td>BB071548</td> <td>BB071548 BB071548</td>	BB071548	BB071548 BB071548

## ALIGNMENTS

RESULT 1  
AZ296526/c  
LOCUS  
DEFINITION  
PC1-23-160G18.TV PC1-23 Mus musculus genomic clone PC1-23-160G18  
' genomic survey sequence.  
ACCESSION  
AZ296526  
VERSION  
AZ296526.1 GI:9538311  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatman, S., Akintet  
, B., Levine, M., McGann, S., Teegaye, G., Geer, K., Krol, M., de Jong, P.  
and Fraser, C. M.  
Mouse BAC End Sequences from Library PC1-23  
Unpublished  
Other GSSs: PC1-23-160G18.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: szhao@tigr.org  
Clones are derived from the mouse BAC library PC1-23. For BAC

FEATURES	Location/Qualifiers
source	1. .593

Query Match	6.3%	Score 92;	DB 28;	Length 593;
Best Local Similarity	100.0%;	Pred. No. 1.1e-34;		
Matches 92;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	Db	Qy	Db
762	GCGCTGAGATGCGCTGCGGAGTCTCGGACCTTCCGGATGAGCGCCAACTTCAAG	821	
250	GCGCTGAGATGCGCTGCGGAGTCTCGGACCTTCCGGATGAGCGCCAACTTCAAG	191	
822	ATCGAGACCTTGGCGCTTGGCCCAACTATCAT	853	
190	ATCGAGACCTTGGCGCTTGGCCCAACTATCAT	159	

RESULT 2	LOCUS	DEFINITION
BY708009	1025 bp	mRNA
BY708009	full length	linear
BY708009	RIKEN	adult male small intestine Mus
musculus cDNA clone 2010001M19 5', mRNA sequence.		

ACCESSION	BY708009
VERSION	BY708009.1
	GI:27119192

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

REFERENCE  
AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamana, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Haesaga, Y., Nogami, A., Schonbach, C., Gobjoiri, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schirni, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chochia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Gilmmond, S., Gustinch, S., Hirokawa, N., Jackson, J. J., Jarvis, E. D., Kanai, A., Kawai, H., Kawasawa, Y., Kezizertsi, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lemhard, B., Lyons, P. A., Maglott, D. R., Maltsev, L., Marchionni, L., Mckenzie, L., Miki, H., Nagasakima, T., Numata, K., Okido, T., Pavani, M. J., Pereira, G., Pesole, G., Petrovsky, N., Pillai, R., Ponting, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sanderlin, A., Schneider, C., Sempile, C. A., Secon, M., Shinada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verrado, R., Wagner, T., Wahlstedt, C., Wang, Y., Matdabe, Y., Wells, C., Wilming, L. G., Wyshaw-Borris, A., Yanagisawa

TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs	Nature 420, 563-573 (2002)	22354683	12466851	Contact: Yoshihide Hayashiraki, <a href="mailto:yoshihide@mol.f.u-tokyo.ac.jp">yoshihide@mol.f.u-tokyo.ac.jp</a>

M. Yang, I., Yang, L., Yuan, Z., Zou, L., Zhao, M., Zhu, Y., Zimmer, A., Carninci, P., Hayashizaki, M., Hirozane-Kitashima, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Aikawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, T., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yamanishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshinide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/

Adachi, Y., Aizawa, K., Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murate, M., Nakamura, M., Nomura, K., Numasaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sato, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

**FEATURES**  
**SOURCE**

```

location/Qualifiers
1..1025
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="201001M19"
/sex="male"
/tissue_type="small intestine"
/dev_stage="adult"
/lab_host="SOLR"
/clone.lib="Riken full-length enriched, adult male small intestine"
/note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGACGCCGCCAACAATTCGAGTTTTCCTTTTTCCTTTCVN 3'], cDNA was prepared by using triethanol thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGACGCCGCCAACAATTAATTCGAGTTAATTAATTAATCCCCCCCC

```



```

BASE COUNT      306 a      485 c      397 g      352 t
ORIGIN
Query Match      4.1%; Score 60; DB 11; Length 1540;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      762 GCGCGATGCGCGCGCGTCTCTGCCACCTTCCGAGATGACCCAACTTACAAG 821
      |||||||
Db      544 GCGCGATGCGCGCGCGTCTCTGCCACCTTCCGAGATGACCCAACTTACAAG 603

RESULT 4
BX445903      923 bp      mRNA      linear      EST 15-MAY-2003
LOCUS      BX445903 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION      CSDDF015YD06 5-PRIME, mRNA sequence.
ACCESSION      BX445903
VERSION      BX445903.1 GI:30774371
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 923)
AUTHORS      Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2626.r
Contact : Feng Liang Email : fliang@life-tech.com URL :
http://fulllength.invitrogen.com/invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS2BAX82A08_AX142B4_1.
FEATURES
source
1..923
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSDDF015YD06"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo (dt) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
vector. Library was not normalized."
BASE COUNT      167 a      290 c      214 g      251 t      1 others
ORIGIN
Query Match      1.6%; Score 24; DB 13; Length 923;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      820 AGATGAGACCTGCGCTTGCC 843
      |||||||
Db      18 AGATGAGACCTGCGCTTGCC 41

RESULT 5
BX524720      standard; RNA; EST; 504 BP.
ID      BX524720
XX      AC      BX524720;
XX      SV      BX524720.1
XX      DT      27-MAY-2003 (Rel. 75; Created)
DT      27-MAY-2003 (Rel. 75; Last updated, Version 1)

```

```

XX      RZPD Mus musculus cDNA clone IMAGE9380187204 = IMAGE:2922473 5' EST.
DE      EST, expressed sequence tag.
XX      Mus musculus (house mouse)
XX      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OS      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
XX      [1]
XX      1-504
XX      RA      Heil O., Ebert L., Neubert P., Peters M., Radlof U., Schneider D.,
RA      Korn B.;
RT      Submitted (28-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL      RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer
RL      Feld 580, D-69120 Heidelberg, Germany
XX      RZPD; IMAGE9380187204.
XX      RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
CC      Mouse Unigeneset - RZPD2 (RZPDLIB No.981)
CC      http://www.rzpd.de/cloneCards/cgi-bin/showlib.pl.cgi/response?libNo=981
CC      Contact: Ina Rofis
CC      RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
CC      Heubnerweg 6, D-14059 Berlin, Germany
CC      Tel: +49 30 32639 101
CC      Fax: +49 30 32639 111
CC      www.rzpd.de
CC      This clone is available royalty-free from RZPD;
CC      contact RZPD (clone@rzpd.de) for further information.
CC      Seq primer: T7, Primer sequence: TAATACGACCTCATATAGCG
XX      Key      Location/Qualifiers
FH      source
FH      1..504
FT      /db_xref="taxon:10090"
FT      /note="Cloned unidirectionally. Primer: Oligo dt.
FT      size-selected 1.5 kb for average insert size 2 kb. Primary
FT      library; non-amplified. This library was constructed by M.
FT      Miene (University College London, Dept of Anatomy and
FT      Developmental Biology)."
FT      /organism="Mus musculus"
FT      /clone="IMAGE9380187204"
FT      /clone_lib="Mouse mouse WT8"
FT      /dev_stage="13.5 dpc"
FT      /lab_host="DH12S"
SQ      Sequence 504 BP; 90 A; 162 C; 192 G; 60 T; 0 other;
XX      XX
Query Match      1.6%; Score 23; DB 4; Length 504;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      732 GAGCGAACCGCATGCACCACT 754
      |||||||
Db      471 GAGCGAACCGCATGCACCACT 493

RESULT 6
BU053975      517 bp      mRNA      linear      EST 26-AUG-2002
LOCUS      BU053975
DEFINITION      UT-M-PD0-bzg-b-11-0-UI_r1 NIH_BMAP_FDO Mus musculus cDNA clone
IMAGE:6403258 5', mRNA sequence.
ACCESSION      BU053975
VERSION      BU053975.1 GI:22494052
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      NIH-MGC http://nigc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE

```



JOURNAL  
COMMENT

Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/INLNL at:  
http://image.lnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers

FEATURES  
SOURCE

1. .517  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6403258"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 12.5 dpc"  
/lab\_host="MDH10B (T1 phage resistant)"  
/clone\_lib="NIH\_BMAP\_FDO"  
/note="Organ: brain; Vector: pYX-Asc; Site: 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is TGAAGAGCC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). Gene Discovery in the Developing Mouse Nervous  
System", supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT  
ORIGIN

90 a 166 c 194 g 67 t

Query Match 1.6%; Score 23; DB 13; Length 517;  
Best Local Similarity 100.0%; Pred. No. 4.8;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754  
|||||  
Db 449 GAGCGCAACCGCATGCACAACCT 471

RESULT 7  
BG808248 600 bp mRNA linear EST 20-DEC-2001  
LOCUS 2083-52 Mouse B14.5 retina lambda ZAP II library Mus musculus cDNA,  
DEFINITION mRNA sequence.  
ACCESSION BG808248  
VERSION BG808248.1 GI:17955225  
KEYWORDS EST.  
SOURCE Mus musculus  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE 1 (bases 1 to 600)  
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L., and Klein, W.H.  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL MEDLINE 21671825  
PUBMED 11812828  
Contact: Klein WH  
Department of Biochemistry and Molecular Biology

University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers

FEATURES  
SOURCE

1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse B14.5 retina lambda ZAP II library"  
94 a 238 c 161 g 107 t

BASE COUNT  
ORIGIN

Query Match 1.6%; Score 23; DB 12; Length 600;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754  
|||||  
Db 4 GAGCGCAACCGCATGCACAACCT 26

RESULT 8  
BU924937 600 bp mRNA linear EST 30-OCT-2002  
LOCUS 7103-91 Mouse B14.5 retina lambda ZAP II library Mus musculus cDNA,  
DEFINITION mRNA sequence.  
ACCESSION BU924937  
VERSION BU924937.1 GI:24428820  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE 1 (bases 1 to 600)  
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,  
White, R.A., Beremand, P.D., Thomas, T.L., Gan, L., and Klein, W.H.  
Gene expression in the developing mouse retina by EST sequencing  
and microarray analysis  
Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL MEDLINE 21671825  
PUBMED 11812828  
Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.  
Location/Qualifiers

FEATURES  
SOURCE 1. .600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse B14.5 retina lambda ZAP II library"  
99 a 194 c 222 g 82 t 3 others

BASE COUNT  
ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 600;  
Best Local Similarity 100.0%; Pred. No. 4.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACCT 754  
|||||  
Db 363 GAGCGCAACCGCATGCACAACCT 385

RESULT 9  
BU709606 673 bp mRNA linear EST 26-NOV-2002  
LOCUS BU709606

```

DEFINITION      UI-M-FRO-cbe-g-01-0-UI.r1 NIH BMP_FRO Mus musculus cDNA clone
IMAGE: 6808442 5', mRNA sequence.
ACCESSION       BU709606
VERSION         BU709606.1 GI:23643245
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 673)
REFERENCE       NIH-MGC http://mgc.nci.nih.gov/.
               National Institutes of Health, Mammalian Gene Collection (MGC)
               Unpublished
COMMENT        Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-remail.nih.gov
               Tissue Procurement: Dr. Jim Lin, University of Iowa
               cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
               DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
               DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
               Clone Distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LNLN at:
               http://image.llnl.gov
               This clone was contributed by the Brain Molecular Anatomy Project
               (BMAP)
FEATURES        Seq primer: PYX-5.
SOURCE          Location/Qualifiers
               1..673
                   /organism="Mus musculus"
                   /mol_type="mRNA"
                   /strain="C57BL/6"
                   /db_xref="taxon:10090"
                   /clone="IMAGE: 6808442"
                   /tissue_type="whole brain"
                   /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
                   /lab_host="DH10B (T1 phage resistant)"
                   /clone_1ib="NIH BMAP FRO"
                   /note="Organ: Brain; Vector: pyX-Asc; Site:1: EcotR I;
                   Site:2: Not I; The library was constructed according
                   Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
                   1996. Denatured RNA was size fractionated on a 1% agarose
                   gel. First strand cDNA synthesis was primed with oligo-dT
                   primer containing a Not I site. Double strand cDNA was
                   size selected according to mRNA size fraction, ligated
                   with EcotR I adaptor, digested with NotI and then cloned
                   directionally into pyX-Asc vector. The library tag
                   sequence located between the Not I site and the polyA tail
                   is AGCGACAGAG. This library was created for the University
                   Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                   Developing Mouse Nervous System', supported by National
                   Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
                   program coordinator."
BASE COUNT      122 a 219 c 250 g 81 t 1 others
ORIGIN
Query Match     1.6%; Score 23; DB 13; Length 673;
Best Local Similarity 100.0%; Pred.No. 5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
Oy              732 GAGCGCAACCGCATGCACAACT 754
                |||||||
                |||||||
Db              570 GAGCGCAACCGCATGCACAACT 592
RESULT 10
CA945402
LOCUS          CA945402 687 bp mRNA linear EST 30-DEC-2001
DEFINITION    UI-M-FRO-cdh-1-12-0-UI.r1 NIH BMP_FRO Mus musculus cDNA clone
IMAGE: 6828925 5', mRNA sequence.
ACCESSION     CA945402
VERSION       CA945402.1 GI:27433882
KEYWORDS      EST.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus

```

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 687)	
TITLE	NIH-MGC http://mgc.nci.nih.gov/.	
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: csapbs-remail.nih.gov Tissue Procurement: Dr. James Lin, University of Iowa CDNA Library preparation: Dr. M. Bento Soares, University of Iowa DNA sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNL at: http://image.lnl.gov This clone was contributed by the Brain Molecular Anatomy Project (BMAP)	
FEATURES	Seq primer: PYX-5. Location/Qualifiers 1..687 /organism="Mus musculus" /mol_type="mRNA" /db_xref="taxon:10090" /clone="IMAGE: 6828925" /tissue_type="whole brain" /dev_stage="embryo 12.5 dpc" /lab_host="DH10B (TI phage resistant)" /clone_id="NIH_BMAP_F00" /note="Organ: brain; Vector: PYX-Asc; Site: 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldi, Lemon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with Not I, and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is TGAGGAGGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."	
BASE COUNT	117 a 228 c 250 g 91 t 1 others	
ORIGIN		
Query Match	1.6%; Score 23; DB 14; Length 687;	
Best Local Similarity	100.0%; Pident. No. 5;	
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	732 GAGCGCAACCGCATGCACACCT 754       	
DB	449 GAGCGCAACCGCATGCACACCT 471	
RESULT 11	B0058877 689 bp mRNA linear EST 26-AUG-2002	
B0058877	UI-M-FRO-ca-k-23-0-UI.r1 NIH_BMAP_FRO Mus musculus CDNA clone	
LOCUS	IMAGE:6413710 5', mRNA sequence.	
DEFINITION		
ACCESSION	B0058877	
VERSION	B0058877.1 GI:22499166	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 689)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/.	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished	
COMMENT	Contact: Robert Strausberg, Ph.D.	

Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: PYX-5.

#### FEATURES

Location/Qualifiers

1. .689  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6413710"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_id="NIH\_BMAP\_FR0"  
/note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 128 a 228 c 255 g 78 t  
ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 689;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754  
DB 626 GAGCGCAACCGCATGCACAACT 648

RESULT 12  
LOCUS BU612495 730 bp mRNA linear EST 20-FEB-2003  
DEFINITION UI-M-FR0-cbc-k-21-0-UI.r1 NIH\_BMAP\_FR0 Mus musculus cDNA clone  
ACCESSION BU612495  
VERSION BU612495.1 GI:23278710  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 730)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be obtained from Dr. M. Bento Soares, [bento-soares@iowa.edu](mailto:bento-soares@iowa.edu)

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
Seq primer: PYX-5.  
Location/Qualifiers

1. .730  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="UI-M-FR0-cbc-k-21-0-UI"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_id="NIH\_BMAP\_FR0"  
/note="Organ: Brain; Vector: PYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 131 a 244 c 264 g 91 t  
ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 730;  
Best Local Similarity 100.0%; Pred. No. 5;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754  
DB 570 GAGCGCAACCGCATGCACAACT 592

RESULT 13  
LOCUS BQ178789 814 bp mRNA linear EST 30-APR-2002  
DEFINITION UI-M-EV0-bwt-k-07-0-UI.r1 NIH\_BMAP\_EV0 Mus musculus cDNA clone  
ACCESSION BQ178789  
VERSION BQ178789.1 GI:20354281  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 814)  
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: PYX-5.

FEATURES  
Location/Qualifiers  
1. .814  
/organism="Mus musculus"  
/mol\_type="mRNA"

```

/strain="C57BL/6"
/db xref="taxon:10090"
/clone="IMAGE:5701758"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1lb="NIH BMAP EV0"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail 'is GTGCGTGGAA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT      151 a      268 c      291 g      102 t      2 others
ORIGIN

Query Match      1.6%; Score 23; DB 13; Length 814;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      732 GAGCGCAACCGCATGCACACCT 754
        |||||||||||||||||||||||
DB      663 GAGCGCAACCGCATGCACACCT 685

RESULT 14
LOCUS      BU054481      823 bp      mRNA      linear      EST 26-AUG-2002
DEFINITION      UI-M-FD0-bzj-1-24-0-UI.r1 NIH BMAP_FD0 Mus musculus cDNA clone
IMAGE:6404447 5', mRNA sequence.
ACCESSION      BU054481
VERSION      BU054481.1
KEYWORDS      GI:22494558
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 823)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.
FEATURES
source      Location/Qualifiers
1..823
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6404447"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_1lb="NIH_BMAP_FD0"

```

```

/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail 'is TGAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT      155 a      273 c      289 g      104 t      2 others
ORIGIN

Query Match      1.6%; Score 23; DB 13; Length 823;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      732 GAGCGCAACCGCATGCACACCT 754
        |||||||||||||||||||||||
DB      672 GAGCGCAACCGCATGCACACCT 694

RESULT 15
LOCUS      CA979119      932 bp      mRNA      linear      EST 06-JAN-2003
DEFINITION      AGNCOURT_11295215 NIH_MGC_164 Mus musculus cDNA clone
IMAGE:30146192 5', mRNA sequence.
ACCESSION      CA979119
VERSION      CA979119.1
KEYWORDS      GI:27511773
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 932)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
Plate: NDAM0061 row: h column: 09
High quality sequence start: 16
High quality sequence stop: 640.
FEATURES
source      Location/Qualifiers
1..932
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30146192"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH_MGC_164"
/notes="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2: NotI; Non-normalized full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) cloned directionally, priming method: Oligo-dT. cDNA enrichment: >1k bp. Average insert size 1.8k bp. Priming sequence: 5'GACTGCTTACATCCGAGCGGCCGCC(T) 3'. Tissue contributed by: David Rowe. Library constructed by Resgen, Invitrogen Corp."

BASE COUNT      153 a      349 c      276 g      153 t      1 others

```

ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 932;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754  
|||  
Db 561 GAGCGCAACCGCATGCACACCT 583

Search completed: January 29, 2004, 22:18:31  
Job time : 2908 secs

**THIS PAGE BLANK (USPTO)**